Laboratories for Population Statistics
Manual Series No. 2
August 1972

OPERATIONAL CONTROL OF SAMPLE SURVEYS

10984

10984

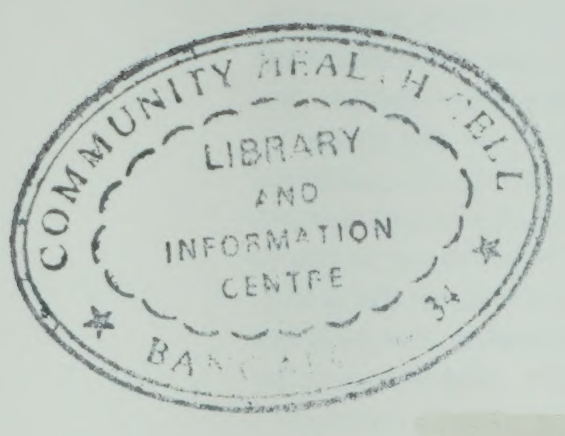
El Programa Internacional de los Laboratorios Estadísticos de la Población, se dedica a la investigación de mejores métodos para la medida de los cambios ocurridos en la población y al análisis de datos demográficos. Para este fin, el Programa espera establecer y mantener "laboratorios" experimentales en los cuales nueva metodología pueda ser comprobada y donde aplicable técnicas innovadas para et análisis de datos ya obtenidos puedan ser desarrolladas. Los proyectos se basan en acuerdos cooperativos entre la Universidad de Carolina del Norte, en Chapel Hill, é instituciones científicas de varios países. El Programa se lleva a cabo gracias a la participación de estas instituciones y a la asistencia de la Agencia para el Desarrollo Internacional del Gobierno de los Estados Unidos. Publicaciones científicas y técnicas son distribuídas, sin costo alguno, a un número limitado de instituciones y personas interesadas en problemas relacionados con medidas demográficas de la población. Las diferentes opiniones y conclusiones pertenecen a los autores respectivos, por lo que no deben ser atribuídas a las instituciones patrocinadoras.

Le Programme international des laboratoires statistiques de population s'adresse au perfectionnement des procédés qui mesurent les variations dans la population et à l'analyse des données démographiques. C'est avec ce but que le programme cherche à établir et à maintenir des "laboratoires" experimentals où de nouvelles métodes démographiques peuvent être verifiées et où applicable, des voies innovatrices d'analyse des données existantes seront développées. Les projets sont basés sur des accords coopératifs entre l'Université de la Caroline du Nord à Chapel Hill et des institutions scientifiques en divers pays. Ce programme se réalise grâce à la participation de ces institutions et l'assistance de l'Agence pour le Développement International du Gouvernement des Etats-Unis. Les publications scientifiques et techniques sont distribuées gratuitement aux institutions et aux individus qui traitent des problèmes démographiques. Les opinions et les conclusions exprimées sont celles des auteurs et non celles des institutions nommées ci-dessus.

The International Program of Laboratories for Population Statistics is dedicated to research on improved methods
change and to the analysis of demograp
gram seeks to establish and support exp
measurement methodology can be teste
sis of existing data can be developed.
agreements between the University of
scientific institutions in various countri
through the participation of these inst
Agency for International Development
States. Scientific and technical issues
number of institutions concerned with
The opinions and conclusions expressed
and should not be attributed to the spo

Community Health Cell
Library and Information Centre
359, "Srinivasa Nilaya"
Jakkasandra 1st Main,
1st Block, Koramangala,
BANGALORE - 560 034.
Ph : 2553 15 18 / 2552 5372
e-mail : chc@sochara.org

Laboratories for Population Statistics
Manual Series No. 2
August 1972



OPERATIONAL CONTROL OF SAMPLE SURVEYS

Walt R. Simmons

International Program of Laboratories for Population Statistics
Forrest E. Linder, Director

The Department of Biostatistics
School of Public Health

The Carolina Population Center

The University of North Carolina
Chapel Hill, N.C. 27514
U.S.A.

◄ à nos lecteurs

Cet article traite d'un aspect important des enquêtes qui est trop souvent neglige ou pas suffisamment mis en lumière: la valeur des opérations entreprises pour réaliser une enquête est étroitement liée à l'établissement du plan de cette enquête. Au cours des dernières décades, l'éventail des méthodes de sondage probabiliste s'est étendu de façon considérable entraînant un accroissement des possibilités des chercheurs et des statisticiens, pour découvrir des faits et fournir des informations adéquates touchant à un grand nombre de sujets relatifs aux phénomènes sociaux, démographiques et économiques. Mais, ces possibilités sont effectives seulement quand la conduite de l'enquête conforme au plan qui en avait été établi. La concordance entre l'exécution et le plan de l'enquête ne peut se supposer et naturellement sera réalisée seulement si des actions y tendant sont entreprises.

L'auteur propose des méthodes pour établir des systèmes de contrôle et des techniques permettant d'assurer la surveillance des opérations et de là, la fidélité jugée indispensable; on attache une attention toute particulière à l'appréciation du plan, au contrôle du recueil des donées, à la façon de les transcrire et de les exploiter, à l'élaboration des estimations et enfin, à la préparation du rapport final. On donne un grand éventail de procédés théoriques et aussi de méthodes empiriques. Toutes ces méthodes ne doivent pas être appliquées systématiquement à toutes les enquêtes. Pour une enquête de petite envergure, les modalités de contrôle peuvent être simples. Pour une enquête plus étendue et plus importante, les mécanismes de contrôles devront être plus nombreux et plus poussés. Ce manuel a pour objet de dégager les principaux aspects de l'enquête qui devront être contrôlé et de proposer les outils de contrôle appropriés. Il a aussi pour objet d'encourager une certaine habileté propre à prévoir les risques d'erreurs possibles dans le déroulement d'une enquête et de chercher les moyens pour enfermer celles-ci dans des limites raisonnables.

to our readers ▼

Operational Control of Sample Surveys treats a significant aspect of survey activity that is too often neglected or underemphasized. This is the steps that must be taken to assure that actual execution of a survey is faithful to its design. The imaginative development of probability sampling in recent decades has increased enormously the potential of researchers and statisticians to discover facts and produce relevant information on a wide range of social, demographic, and economic matters. But that potential is realized only when conduct of the survey mirrors the design under which it was planned. Concordance between execution and design cannot be assumed, and indeed will exist only if positive actions are taken to see that it does occur.

Operational Control offers guidelines for establishing control systems and

a nuestros lectores ►

Este artículo trata de un aspecto significativo del proceso de las encuestas que frecuentemente no se tiene en cuenta o es escasamente acentuado. Se refiere a los pasos que deben tomarse para asegurar que la ejecución de una encuesta sea el fiel reflejo de su diseño. El ingenioso desarrollo del muestreo probabilístico en las últimas décadas ha incrementado enormemente el potencial de los investigadores y estadísticos para descubrir hechos y obtener información relevante en el amplio campo de los hechos sociales, demográficos y económicos. Pero el potencial se materializa sólo cuando la ejecución de la encuesta se refleja en el diseño bajo el cual se planea. La concordancia entre ejecución y diseño no puede ser supuesta, y en realidad solo existirá si se llevan a cabo ciertas acciones dirigidas a obtener tal concordancia.

El autor ofrece guías para establecer sistemas y técnicas de control para mantener vigilancia sobre las operaciones, y por lo tanto, proporciona la seguridad nacesaria. Se da atanción a la validez del diseño, al control sobre la recolección de información, a la edición y al procesamiento, y a la preparación de los estimadores e informe final. Se presenta una variedad considerable de procedimientos formales y de artificios de control subjetivo. No todos estos métodos deben ser utilizados en cado encuesta. Para un proyecto pequeño, la estructura de control puede ser modesta. Para una encuesta mayor y más importante, los mecanismos de control deben ser más extensos y compresivos. El estudio tiene por objetivo identificar aspectos claves de operación que deben ser controlados y sugerir herramientas de control apropiadas; pero aun más, se propone fomentar un hábito de pensamiento que reconozca el riesgo de error en cualquier procedimiento de encuesta y que busque los medios de mantener tal riesgo dentro de limites aceptables.

techniques for maintaining surveillance over operations, and thus providing the needed assurance. Attention is given to validation of the design; to control over data collection, editing, and processing; and to the preparation of estimates and the final report. A considerable variety of both formal procedures and more subjective control devices is presented. Not every one of these methods should be used in every survey. For a minor undertaking, the control structure can be modest. For a larger, more important survey, the control mechanisms should be more extensive and comprehensive. The objective of the manual is to identify key aspects of operation which should be controlled, and to suggest appropriate control tools; but more than that, to encourage a habit of thinking that recognizes risk of error in any survey procedure, and seeks means of keeping that risk within tolerable bounds.

Table of Contents

1. Introduction 5

2. The Principle of Control 5

3. Types of Survey 6

4. Outline of the Operational Control System 7

   4.1 Reconciliation of Design and Objectives 7
   4.2 Verification of Theoretical Design 7
   4.3 Control over Data Collection 7
   4.4 Control of Data Processing 7
   4.5 Control of Tabulation and Estimation 8
   4.6 Control over Report Manuscript and Publication 8

5. Reconciliation of Design and Objectives 8

   5.1 General Considerations 8
   5.2 An Illustration 9
   5.3 Confirmation of Plan 11
   5.4 A General Reminder 11

6. Validation of Theoretical Design 11

   6.1 Scope of the Second Phase of Control 11
   6.2 Rationale of Design 11
   6.3 Mathematical Validation 13
   6.4 Verification of Arithmetic 14
   6.5 Mesh of Data and Design 14
   6.6 Mean Square Error 14
   6.7 Resources and Logistics 15

7. Control Over Data Collection 16

   7.1 Planning and Operations 16
   7.2 Pretests 16
   7.3 Public Relations 18
   7.4 Recruitment and Training of Agents 18
   7.5 Dress Rehearsal 19

7.6 Systems of Control over Field Collection 20
7.7 Sampling Controls 21
7.8 The Control Subsample 22
7.9 An Observation Program 30
7.10 Error Reports from Central Editing 31
7.11 Combining Control Evidence 31

8. Editing and Processing Controls 32

8.1 Introduction 32
8.2 Document Flow Control 32
8.3 General Editing Procedures 33
8.4 Multiple Coding of Subsamples 35
8.5 Imputation 38

9. Control of Tabulation and Estimation 39

9.1 Character of the Chapter 39
9.2 Transcription Controls 39
9.3 Control Totals 39
9.4 Imputation Procedures 39
9.5 Validation of Programming 40
9.6 Special Gross-Tabulation Checks 41
9.7 Comparative Analysis 41

10. Control Over Report Manuscript 42

10.1 Analytic Methods 42
10.2 Transcription Verification 42
10.3 Verification of Computations 42
10.4 Completeness of Manuscript 43
10.5 Proofreading 44

11. Concluding Comment 44

Appendix A. A Brief Note on Acceptance Sampling and Process Control 44

Appendix B. An Illustrative Procedure for Guidance in Selecting and Training Interviewers 47

List of References 49

Operational Control of Sample Surveys

Walt R. Simmons

1. Introduction

The activities of nations, governments, business enterprises, universities, and individual persons are complex, but all undertakings directed by rational men can be described in outline as formulation of objectives, planning to meet those objectives, and execution of the plans. Most administrators are coming to realize that wise and productive social action must be based on extensive and detailed factual information. Increasingly that knowledge is being acquired, or at least sought, through the mechanism of the statistical sample survey. By whatever means initial primary objectives may have been determined, it is true for the sample survey, as for other undertakings, that rational behavior calls for planning and then for carrying out the plans.

The planning activity is responsible for producing the survey design; execution is comprised of a series of activities termed survey operations. As will be made clear in subsequent sections of this manual, survey design and survey operations are (or should be) interactive. Moreover, if the purposes of design are to be fulfilled in operations, those operations must be in reasonable accord with the designers' intentions, just as in building a house actual construction must follow the architect's plans. It may be less obvious to the inexperienced surveyor that resulting correspondence between survey design and survey operations is very likely to be highly disappointing unless positive steps are taken to guarantee that correspondence. The purpose of this monograph is to provide guidance to survey managers in adopting techniques which will give much greater assurance that performance will match intention, even though absolute guarantee is never certain and can only be approached with continuous vigilance. Special emphasis will be placed on the merits of independent verification and replication, taking advantage of the principle that the probability of the same error being made by two independent measurements is the product of probabilities of error for the individual measurements and therefore, much smaller.

WALT R. SIMMONS *is Assistant Director (Research and Scientific Development) at the National Center for Health Statistics in Washington, D. C. and Adjunct Professor at the American University in Washington, D. C.*

2. The Principle of Control

The competent statistician has learned and must never forget that he lives and operates in the shadow of an axiom:

The conduct of a survey that lacks built-in operational controls will not mirror the design.

Several corollaries follow immediately. If the survey is not carefully controlled, money and time will be wasted, and results will be of doubtful and unknown validity. The survey will fail to produce solid information, and may actually yield misinformation. The objective which was hoped for will be lost.

Why does an uncontrolled survey fail? It may not be that planning was weak, or instructions unclear, or that enumerators, coders, and clerks were unskilled or dishonest — although any of these factors and others may be present. Rather it is that all of us at times misunderstand, make mistakes, encounter unforeseen situations, and take actions which we had not earlier intended. (Indeed some survey statisticians say ruefully, "If an unfavorable event can happen, it will.") The survey can be successful then only when procedures are devised which collectively will assure that deviations from the intended design are kept within reasonable bounds. This is the principle of operational control, but how is it secured? That cannot be answered in a word, nor in a few sentences. This manual will identify the points at which controls are most necessary, and offer techniques and procedures for control over operations at those points. The recommended methods have been found effective, but the most valuable message the manual carries is the encouragement it may give the practicing statistician or researcher to develop the habit of subjecting *all* survey operations to verifying or validating procedures to the maximum extent that his imagination and resources will allow.

Some readers may say, "Of course the survey should be carried out as planned. Everyone knows that. Any good director or supervisor will see that his employees do their jobs properly." The task is difficult, though. Good intentions, and vigorous general supervision are not enough. *Structural* controls are required. These are the tools of strong supervision. Skill in their use can change failure into success.

3. Types of Survey

There are many kinds of sample surveys. They differ with respect to subject: for example, they may treat human or animal populations and their characteristics — those which deal with people may concentrate on particular aspects of the population such as births and deaths, health, unemployment, income, hours of work, industrial or occupational attachment; or the universe of interest may be crops planted and harvested; or industrial production and sales; or housing; or domestic and international trade; or a host of other areas. The principal data-collecting instruments vary: interview of persons in their homes, or at their places of work; transcription of written records, such as birth registrations or payroll or tax lists; in some circumstances, mail questionnaires distributed and collected through a postal system are appropriate. Surveys may be continuing over time, or intermittent, or conducted on a single occasion. They may be local or nationwide in scope. If nationwide, they may be centrally directed, or they may comprise integration of regional, district, or local efforts. The overall plan may be a scientific probability design, or it may be an *ad hoc* haphazard collection of data through some convenient administrative mechanism. The undertaking may be a major enterprise involving much staff, equipment, and large sums of money. Or it may be a minor effort carried out by a few people, or even a single person, in a brief period of time. The general purpose may be a legal or administrative determination, or it may be research simply to improve the state of knowledge.

Thus it is that surveys differ tremendously in kind in the dimensions just indicated, and in other ways. Appropriate control systems differ accordingly. Some elements of control should be common to nearly all forms of sample survey endeavor. Other control procedures are peculiar to particular survey procedures.

For convenience in presentation, to facilitate understanding, and in order to offer material that is specific and capable of fairly direct application, this manual treats control of statistical operations in the context of one broad class of sample surveys. If a thorough grasp is acquired of the methods described it is believed that knowledge will permit a statistician to construct modifications which will be suitable for many other classes of survey.

The type of survey which the reader is asked to have in mind as operational control is discussed in the present manual, has these general characteristics.

A. It is a large-scale, continuing or intermittent survey of a human population and its characteristics. Particular emphasis is placed on demographic, sociological, and economic

features, with health and vital events likely receiving prominent attention.

B. The overall plan is a probability design. The sample structure may be relatively simple, or it may be complex, involving stratification, clustering, more than one stage, more than one phase, elaborate estimating equations, and other refinements. But it must possess one key attribute: every unit in the target population must have a known non-zero chance for inclusion in the sample.

C. The main data collection is handled by field agents. At most points in the manual, the agent secures his data through interview, but at some through transcription of records.

D. The survey is centrally directed, although usually supervision of data collection is exercised through several district or field offices.

As the researcher or survey director adapts the recommended control procedures to his particular circumstances, which likely will differ in at least some degree from the model survey which is the background of the manual, he may do well to proceed in accordance with these guidelines: For any phase of the survey operation:

(a) Decide what is desired to be accomplished;

(b) write instructions for doing it;

(c) do it;

(d) devise and execute a checking procedure to insure that it is being done correctly and is completed in accordance with the instructions.

4. Outline of the Operational Control System

Subsequent sections of this manual will deal with six phases or subsystems of an overall operational control system which is intended to assure within the limits of available resources that the ultimate survey product will in fact fulfill the original objectives of the undertaking.

4.1 Reconciliation of Design and Objectives. This first phase is the least specific of the control subsystems. It is at one time the most obviously needed, yet often overlooked, and possibly most important single validation of survey effort. The concept is a simple one. It envisages conversion of general purposes into specific objectives. It accepts tentatively the survey plan which has been prepared. It next assumes that the plan is successfully executed, and looks at the likely hypothetical output. The critical issues are whether that output answers the original questions and satisfies the survey purposes. Often for the initial plan, the answer will be "No it will not." Then the initial plan must be modified, and the process repeated until there is thoughtful agreement that if the adopted plan is faithfully executed, there is reasonable expectation that the survey will accomplish its purpose.

4.2 Verification of Theoretical Design. This manual does not discuss in any detail the mechanics of formal sample or survey design. These fundamental matters are treated in textbooks, other manuals, and the technical appendices of many survey reports. It is assumed here that design work in the general sense has been done competently. Nevertheless there are a number of respects in which the formal design needs to be checked and validated. One needs only recall the tragic comment of the design engineer, witnessing the collapse of a new bridge during opening ceremonies: "*There* is that missing decimal point!"

4.3 Control over Data Collection. It is in this third phase of control that survey management is most on trial, and where success or failure of operations is most likely to be determined. A series of control points and processes must be established, which collectively promote agreement between the character of intended evidence and resulting reported data, and which minimize the introduction of extraneous, incorrect, and dishonest data.

4.4 Control of Data Processing. Building upon a base of the output of data collection, all surveys pass through a phase of converting the raw data into edited, coded evidence suitable for tabulation and analytic manipulation. These steps are production-type processes, similar in some respects to those of industrial manufacturing, where quality-control procedures were first developed. They can be kept within acceptable tolerance levels with the classic quality control techniques. But there are also special types of checking that are particularly suitable for coding operations, and statistical devices for validating a sequence of

processing operations. These will be set forth in this manual.

4.5 Control of Tabulation and Estimation. The essence of tabulation of edited and coded data is straightforward arithmetic. Even for complex sample designs, inflation of sample data to universe estimates, and more elaborate estimation of population parameters can be specified by rigid mathematical formulae. It is reasonable to expect that these processes can be carried out with a high degree of accuracy, but it is also reasonable to imagine that it is possible for there to be mistakes in arithmetic, algebra, statement of instructions, programming (when tabulation is accomplished on an electronic computer or other electric or mechanical equipment), and in the treatment of outliers and other unusual items of data. All these errors are likely to occur. They can be very destructive unless positive provision is made for their control.

4.6 Control over Report Manuscript and Publication. If the survey has been well designed and a powerful set of operational controls has been in effect up to this point of the survey, one can expect that valid data have been produced, and that skillful analysis will lead to a good report. One more control subsystem is necessary. This subsystem has two aspects. One is simply verification of transcription and arithmetic which convert worksheet results to numbers, tables, and charts in the manuscript and final report copy. The other is some scheme of review or independent analysis of judgments and conclusions reached by the authors of the report. The review should include editorial scrutiny for clarity, removal of ambiguity, and attractiveness of presentation. This final control procedure is again, like the first subsystem, somewhat more subjective and less specific than the other four subsystems, but that does not justify its omission. It would be a personal tragedy for the survey director or statistician to have produced sound and relevant data through great and careful effort, only to have the end product misrepresent those findings, or be disregarded because it could not be understood, or was simply too unattractive to be read. The statistician should remember that usually the majority of his clientele will judge his performance — perhaps not fairly — not so much by how he planned or conducted his survey, but by the report he prepares of that survey, its findings, and conclusions.

5. Reconciliation of Design and Objectives

5.1 General Considerations. On too many occasions a survey produces "beautiful" tables, charts and conclusions — and still has little relevance or utility for the purposes for which the statistical project was undertaken.

The remedy for these deficiencies must be applied early in the project. General purposes should be connected to specific objectives. Ideally these should be expressed in a series of written questions which the survey needs to answer. The questions should be quite specific, and they should be expressed in defined terms, and capable of being answered in measurable quantities. Scope of the survey with respect to subject matter, definition of population to be surveyed, geographic coverage, time span, length of reference period, and frequency of enumeration must be clearly stated. Decisions are made concerning whether results are needed for only the total universe as a whole or separately for certain geographic subdivisions and specified domains within the universe. At least general levels of acceptable tolerances or sampling error are set. These steps clarify and put into useful operational format initial requirements for the survey.

Assume that an initial survey plan has been outlined and tentatively accepted. Consider next the expected output of such a survey. Does it match the requirements? The first round of planning likely will not have produced a complete match. The projected survey may be deficient in some respects, or it may contemplate the production of information which is not really needed, and thereby be more expensive than is justified. Appropriate adjustments in the plan are made. Note that the title of this section is "Reconciliation of Design *and* Objectives." It may be that considered study of the situation results in the conclusion that one or more of the originally stated objectives are not really feasible. On balance, the evidence may be that certain desired attributes of the population

are not measurable through any survey techniques which circumstances permit. In that situation, it is the objective which should be modified, so that it is re-expressed in terms that are measurable. This pattern of review, comparison, and modification of plans and objectives is repeated until satisfactory agreement between objective and plan is obtained. The cost of any compromise in objective which may have been necessary is more than offset by the value of having chosen a plan which *can* be carried out, as contrasted with taking a course which is doomed to failure from the very beginning.

5.2 An Illustration. Let us suppose that a government has decided it is necessary to undertake a study of the human population of the Province of West Dakota. Officials say that they wish to know how many people there are, how the size of that population is changing, the health status of the people, and the situation in terms of gainful employment.

Assume that designers have prepared a tentative plan for this survey. They decided that the general vehicle should be a household interview of approximately 10 per cent of the population, selected through a multi-stage area sample of West Dakota. Questionnaires have been constructed, and relevant definitions and instructions written. Preliminary plans have been set for securing necessary resources in personnel, space, equipment, sampling materials, publication facilities, and other supporting commodities. A first formal sample design has been developed. Drafts have been completed of instructions for all phases of the survey, including data collection, processing and tabulation. In short, a comprehensive and competent job of early planning has been done.

In the first phase of operational control, the assumption is made that all production will be carried out as planned. The question is, will the survey then satisfy its objective? This in turn raises such questions as the following, each of which should be explored and resolved before the survey goes into the field and even before much energy is devoted to planning the more detailed and specific control procedures to be applied to other and later phases of the survey.

A. Is the geographic scope of the survey and of West Dakota clearly and unambiguously defined? If boundaries are in any respect uncertain, they must be clarified. In this respect, and in all subsequent checks, a positive effort should be made to bring final definitions and determinations into as close accord as feasible with the desires of the sponsors of the survey — remembering always that realities may dictate modifications of original preferences, to the end that the task attempted is a feasible one.

B. Do the definitions provide for whether every person who may be encountered in the survey is properly included or excluded from the universe of study? Are there classes of persons who belong in the universe but are likely to be missed unless special provision is made for them? Is there a firm decision on whether a *de facto* or a *de jure* concept of population has been adopted? Is the effective date of the count specified in an acceptable fashion?

C. Since "household" is one of the building blocks of the data collection procedure, and since it is a component of the definition of persons to be included in the sample, is the definition of the term consistent with the intended inclusion of persons in the population? For example, if it is intended that West Dakotans who have not lived in the Province for a year be excluded, do household membership rules also exclude such persons?

D. The questionnaire should be reviewed to see if it meets the implications of the following list of questions and the principles which they reflect: (1) Will information be provided on all required subjects? (2) Does the questionnaire ask for the *right* information, in terms that are properly defined, that can be readily understood by respondents, and in sufficient detail for appropriate analysis? (3) Will the resulting data and classification systems yield information not alone for the total universe, but for desired geographic and demographic subuniverses? And (4) Will the survey or known other sources provide necessary and appropriate denominators for any rates or ratios which analysts desire? Items E, F, G, and H immediately below offer specific illustrations of how some of these tests can be applied.

E. The questionnaire provides for recording head count, sex, age, occupation, employment status, and health status for each person in every sample household. Passing over, for the moment, the need for data on events as distinguished from data on current condition or status, do these questions provide all the material needed for meeting the survey mission? In answering this question, it is necessary to go beyond the words used originally by the sponsors and to try to satisfy their intention. Sex and age were not explicitly requested by the sponsors, but they clearly are necessary to any useful analysis of survey data. The review probably would conclude that relation to head of household also is needed, and should be added to the questionnaire.

F. Sponsors asked for the situation in terms of gainful employment and health status of the population. Do employment status and health status in the questionnaire, along with the other items provide the answers? Certainly alone, just by themselves, two such questions would not. Consider first employment status. An appropriate course will depend on the nature of the economy in West Dakota, but very likely it will be necessary for the inquiry to be a good bit more specific — perhaps along the following lines. In addition to formal occupation, there should be questions that will distinguish among housewife, small children, children in schools, and persons too old or unable to work, as well as determination of whether the person is employed or unemployed. Employed is itself capable of a variety of definitions, and the one chosen should match the needs of the sponsors. It may well be that knowledge is desired too of class of worker, i.e., whether a person is self-employed, an employee of a government, or working for a private employer. Is the objective to secure employment status, now, during the past week, usually, or over some other time period?

G. Health Status is an even more complex concept than employment. Probably no set of questions administered by interview can yield a fully satisfactory picture of the health status of a person or population. This is almost certainly an area of the survey in which the original charge will need be modified in order to establish an objective that can be attained. Quite possibly the target may be revised to answer the following questions: (1) Is the person physically impaired, i.e., blind, deaf, missing arms or legs, unable to feed himself, or to move about? (2) Is he able to do the type of work to which he is accustomed, or which persons of his age and sex can usually do? (3) Does he have any of a specific list of chronic diseases? (4) How many days in the past month has he been confined to bed all or most of the day? These questions will not produce a complete description of the health status of the population. But they, or other similar questions, supported by data on deaths and causes of death (to be discussed shortly), can yield valuable evidence on health for the government officials. Since they represent inquiries and objectives that are attainable, they promise less disappointment than would a more amorphous attempt to enumerate health status.

H. The survey is asked to provide information on changes in the size of the population. Presumptively, the survey planners will have converted this quite general request into specifics. Assume they have decided to mount a continuing undertaking, and to repeat the survey at annual intervals. Thus once each year they will enumerate a cross section of the population, and this will provide the basis for annual assessment of net change of the profile by age, sex, and other measured characteristics. Since health and the dynamics of population change are also of interest, they have provided for enumerating births and deaths. (Experience has shown that determination of birth and death rates by an intermittent interview procedure is difficult. But remember, we assume in this section that the chosen plan can be executed.) Review of this part of the plan must assure that procedures are designed to count *all* births and deaths which occurred to persons associated with the sample areas during the reference period (perhaps the year preceding interview). Several valid variants of this formula are acceptable. The point is that definitions and counting must provide for consistent formulation of numerators with denominators of the birth and death rates, while attention is given also to preventing omission of

certain classes of persons or events from either numerator or denominator, or to erroneous inclusions of persons or events. Particular attention must be given to rules for attaching a person uniquely to an area, and to treatment of migration into and out of West Dakota during the reference period. Since the interviews are with the living population, special attention must be given to proper inclusion of deaths, and most especially in instances in which death may have dissolved a household.

It is again emphasized that it is not the purpose of this manual to recommend ideal survey designs, but rather to offer guides for controlling surveys after they have been planned. There is an extensive literature on designs for measuring population change and vital events. See for example references [10,11,12,13].

5.3 Confirmation of Plan. Having reviewed each of the main features of the survey plan in terms of specific objectives, and likely having modified the original requests in some respects, it is desirable to secure the concurrence and approval of sponsors, i.e., the officials who have authorized or commissioned the undertaking, before proceeding further. The statistician or project director may be hesitant to face the sponsor at this stage, having chosen a plan which may offer somewhat less than the sponsor expected, and not yet being absolutely certain that he can ultimately produce even the lesser product. But it is far better to make certain at this point that a fully agreed upon task is being undertaken than to discover at the end of the survey that what he has done is unacceptable. Perhaps any differences of opinion at the early stage can be happily resolved with modest effort. It is quite likely that the sponsors will be pleased to see the researcher proceeding in an orderly, business-like manner, and to learn in fairly specific and concrete form just what kind of evidence and information the survey is likely to produce.

5.4 A General Reminder. From time to time in this manual, for purposes of illustration and in order to focus on a specific point, there will be discussions of a particular example, as in section 5.2 above. It should be remembered that these are examples only. Surveys and survey objectives differ widely. A particular procedure which is appropriate, or even necessary for one survey, may be ill-suited to another. The user of the manual is urged then not only to become acquainted with the specific procedures which are suggested, but even more importantly to seek understanding of principles behind the procedures and of the fundamental purposes of those procedures. For instance, in this first phase of operational control, discussed in section 5, the central purpose is to take positive steps to assure that the chosen strategy for a survey, if successfully executed, will in fact satisfy the objectives of sponsors. It means that the operating statistician starts out with the confidence that if he does his job well, he will be engaged in a productive enterprise. He avoids the risk that even if he should do a good technical job, the resulting data might be unsatisfactory, or even serve no useful purpose, and might be of no interest to any significant consumer.

6. Validation of Theoretical Design

6.1 Scope of the Second Phase of Control. From this point on it is assumed that a general strategy has been agreed upon, following thoughtful review, and final confirmation by sponsors and administrators as well as by the project manager. Indeed a reasonably complete design for the survey has been laid out. At this point, a second control phase should be introduced before data collecting operations begin. This control is technical in nature, in the sense that it is intended to validate the work of the mathematicians and survey statisticians who are the professional survey designers and planners.

It should be made particularly clear at this point that the introduction of checks and validating procedures for the chosen design is not in any sense a reflection on the wisdom or skill of the designing statistician. Rather it is recognition of our sound guiding principles that the risk of error is always present, but can be minimized through replicating and control processes. The true professional statistician will welcome and even insist that his design work be validated.

6.2 Rationale of Design. Remember an assumption of this manual is that a generally

sound overall design has been formulated. It has now been accepted. The control objective in this phase is not to search anew for the best possible design, but rather to review the chosen design, to look for flaws in details, perhaps to introduce modifying improvements, in order to augment confidence that the blueprint for the survey does indeed offer a workable pattern, and thus to confirm that the design can reasonably be expected to be successfully executed.

If circumstances permit, the ideal course is to obtain an independent review of the total plan by one or more competent statisticians who are knowledgeable of the survey objectives, but who were not the responsible original designers. The principle which directs this course is that two or more largely independent judgments are less likely to make the same mistakes, or overlook the same shortcomings. If independent reviews are not available, the task is more difficult, and more dangerous, but valuable review is still possible. The same statistician on a second trial can uncover weaknesses in an earlier effort if he consciously approaches his task with a critical and questioning frame of mind. He can deliberately construct alternatives, compare the consequences, and choose the better route. Often he can arrange for associates or subordinates to check particular aspects of the design.

One fruitful technique is to consider the impact which unexpected circumstances or population parameters would have if they should be encountered in conduct of the survey. For example, the design may have been predicated on speculation that the population coefficient of variability of a key item is 40 per cent within primary sampling units (P.S.U's) and 10 per cent between P.S.U's. Construction of a table of sampling errors which are functions of within- and between-variances over considerable ranges from the speculated values will identify areas of allocation of sampling resources to within- and between-P.S.U's in which the design is most stable or robust. This exercise may confirm the original allocations or may suggest a modification that promises better insurance against the unexpected.

In some instances the review may suggest fortifying the initial design, by introducing an element of flexibility into the sampling or estimating plan. For example, scarcity of data may have led to an uncertain determination of sample size or of unit costs. It may be feasible to break the parent sample into, say, four random subsamples or panels, and then schedule operations so that preliminary evidence from the first panel can be used to decide whether three or all four panels must be used to secure needed precision or to stay within available budget. Or perhaps the estimating scheme specified the use of a ratio estimate, expecting to take advantage of correlation between two variables, x and y. It may be wise to modify the plan so that the ultimate decision on type of estimate will depend upon evidence of the degree of correlation which materializes in the actual survey data.

Suppose the survey seeks to measure the birth rate by collecting interviews in a sample of land segments. The estimating scheme makes certain assumptions about the characteristics of persons who migrate into and out of those segments during the reference period. These assumptions are critical to final results and should be reviewed very carefully. With only modest modification of field procedures and the addition of a few questions in the interviews, it often will be possible to build into the survey techniques for evaluating the migration assumption, and possibly strengthening the estimates of birth rate.

We do not treat in any detail at this point the issues of relevance, specificity, clarity, and answerability in the questionnaire, assuming that they either have been covered in phase I control, or will be resolved in pretests and pilot studies to be treated subsequently. But the reviewer must necessarily keep in mind that the real survey, as distinct from a model on the drafting board, will be dealing with responses in terms of numbers, replies of "yes" or "no", names of people and places, and classifying attributes rather than the more tractable x's and y's of his equations. Accordingly, he should try to assure himself that there will be congruence between the collected data and the math-

ematical models developed in the planning phase.

The design should be checked for presence and reasonableness of a number of provisions. Some of these are discussed in subsequent sections. A supplementary five-point check list is noted at this point.

A. Is there an adequate plan for handling non-response? First, there need be standards set for maximum allowable non-response, both in failure to respond by the entire sampling unit, and in omission of some items of data for sampling units which report incompletely. Second, a continuing surveillance procedure should be established, which, along with corrective measures, will assure that the standards are met.

B. It must be expected that both environmental circumstances and the economics of operation are such that not all non-response can be prevented. Therefore, procedures must be established for imputing values for the missing information into survey averages and totals. Remember that imputation, either explicit or implicit, cannot be avoided. It is not a question of *whether* to impute for missing data, but only *how* to impute.

C. Is the contemplated computational procedure arithmetically sound? Perhaps not all possible contingencies can be foreseen at this stage, but one can check to see if plans provide for carrying appropriate numbers of significant digits, recording of needed sub-totals, specifying the handling of blanks and "impossible" items of input, avoiding steps which might involve division by zero, and making sure that the size of expected sums and products is within the capacity of chosen computational equipment.

D. Will the processing and tabulating scheme provide an output *record* of statistics on important editing, imputation, and adjustment steps? For example, the bookkeeping should provide a record of the number of unit non-responses, numbers of missing items, the volume of key imputations and the sizes of ratio adjustments. This record is invaluable to later evaluation of output. It can save the statistician from drawing unjustified conclusions. If the record showed that it had been necessary to impute, say, 60 per cent of the data for a particular item, the statistician would know that the survey had produced no reliable data at all on that item.

E. Does the plan provide a comprehensive and detailed blueprint of document flow and the transmission of information (data) throughout all phases of operation? The working pathway from respondent to publication should be clearly delineated. This matter is discussed further in section 8.2.

F. Is the total procedure adequately documented? There should exist at least one complete "brass-bound file" or master account, which describes in full detail all aspects of the design and associated procedures. This is necessary to protect against faulty memory on the part of designers, as well as against the possible death or absence of those designers at a later stage, when it will often be necessary to know precisely what the plan of the survey or its operation was on a critical point. It also is the foundation for preparation of instruction manuals, which should be written for all phases of the survey operation, and it is the authority of last resort for the analysts who finally draw conclusions and publish the survey results.

6.3 Mathematical Validation. It is recommended that a second statistician check the mathematical and statistical logic, and the algebra of the entire theoretical formal design. Again it is observed that such a step is no reflection on the ability of the principal designers, but only a reasonable precaution. All of us make mistakes. Even the most skillful and experienced person will occasionally forget to square an expression, will use a variance where he should have used a rel-variance, neglect a remainder in a series expansion when the remainder is not trivial, mishandle the algebra in calculating an expected value, use an improper variance formulae, or err in many other ways. If the error is undetected, the entire survey structure may rest on a foundation which contains a serious or even fatal flaw.

Often, in addition to straightforward dependent verification of the original mathematics, it is both desirable and feasible to secure a partially independent derivation, proceeding from slightly different assumptions and arriving at approximately comparable conclusions. When

this is possible, it provides the best kind of insurance.

6.4 Verification of Arithmetic. This step is an adjunct of the one just described, is taken for the same reasons, and is even more obviously necessary. All arithmetic and numerical calculations in the designing and planning phases must be verified on a 100 per cent basis. Several operational hints may be helpful:

A. Particular attention should be focused on location of all decimal points.

B. The *scale* of all results, both intermediate and final, should be considered for reasonableness. Often a supplementary approximating calculation using only two or three significant digits and powers of 10 will produce a confirming figure for comparison.

C. Often, for a complex plan — perhaps a two-stage, stratified, and post-stratified design — it is useful to make parallel computations for a simpler design, and to compare the two results. The tactic here is of course to seek confirmation of a plausible relationship between results from the extended calculations for the complex design and those from the simpler arithmetic where there is less chance for error.

D. Some of the input for these calculations is likely to be speculative in character, and almost certainly to have been transcribed from outside and possibly alternative sources. A careful and questioning review of the input numbers is warranted.

E. Perhaps it is looking for trouble to suggest the possibility that a desk calculator used in making calculations may not be functioning properly. Yet this can happen, and of course one should be reasonably sure that it does not. It is easier to believe in possible equipment malfunction, but harder to avoid, if part of the planning computations are carried out on punchcard or electronic computer machines. In such instances, error can have its source in either the hardware or in its programming. When this type of calculation is involved, it is particularly important to seek confirmation of key results through simpler approximating techniques.

6.5 Mesh of Data and Design. Attention was called earlier in the discussion of review of the overall rationale of the design, to the need for correspondence between collected data and tabulating plans. As the details of tabulation and estimation are finalized, a firm matrix must be established which translates unequivocally each input datum into the appropriate variable or constant of the design pattern. This matrix is a two-way guide. It pictures as well the *source* of data for each variable or constant — each *x, y, a, b, c* — of the formal design. An affirmative answer is required for every question: "Does the plan assure that needed data will be forthcoming for this factor in the design and tabulation plan?"

6.6 Mean Square Error. One wishes it could be said that every survey plan must include a scheme for determining the mean square error of the principal survey results. Mean square error (MSE) is defined as the sum of the variance and the square of the bias of a statistic. The bias is defined as the difference between the expected value of the statistic and the true value of the parameter which the statistic estimates. Variance is the expected value of the square of the difference between the observed statistic and the expected value of that statistic. Total variance includes both sampling variance and measurement variance, but does not include any bias, whether it may arise from sampling or measurement.

Unfortunately the state of the art does not permit the promulgation of a requirement that MSE must be specified. But the art and good practice do justify a demand that the survey plan provide for means of discovering the sampling variance of key statistics, and that the statistician take such steps as he is able to supply evidence on measurement variance and on bias. No plan is complete without such provision.

Other manuals and sampling textbooks, as well as a fairly extensive list of articles in professional journals offer methodology for securing sampling variances and for approaching evaluation of mean square error. This manual offers no recommendations on how the evaluations are attempted, but does insist that the survey plan deal with the topic, and as a minimum provide for determination of sampling variance. It also suggests the desirability of the following provision which is not often found in textbooks. We refer to the following situa-

tion. Frequently the calculation of a precise sampling error is an involved computation which is expensive, time-consuming, and given a low priority in operational demand on personnel and equipment. Too often, then, other tabulations are completed, analysis performed, and even published before the sampling errors have been computed. To protect against this contingency, it may be wise for the survey plan to include a back-up procedure which is a less costly approximating technique for estimating sampling variance. That weaker, but still useful, technique can be utilized in the event that the more fully adequate process does in fact prove to be either unattainable or late in coming forth.

6.7 Resources and Logistics. Every survey plan is predicated upon, or assumes the existence of, the wherewithal to carry out the plan. This means availability of financial resources, manpower of several kinds, space, equipment, materials, and supplies, along with an administrative and operational structure with the capacity to deliver proper and timely logistical support at all phases of the survey.

Examination of the provision for these matters calls for the talents of both the statistician and administrative aides who are knowledgeable of relevant personnel rules, procurement channels and regulations, budgetary procedures, and other management paths. In these areas as in others, the specifics vary with circumstances, but typically the statistician should review the specifications for numbers and kinds of personnel required, and check over the availability of needed equipment and materials, including especially such items as lists of units to be sampled, maps, classification systems — both the systems themselves, and the related coding manuals — transportation and communication devices, file cabinets, duplicating machines, transcribing and tabulating machinery, typewriters, and the long list of small supplies such as paper, pencils, staplers, envelopes, punchcards, index tabs, and file folders. He will assure himself that the survey will in fact have access to any auxiliary data — for instance, census files, or tax and licensing records — which have been incorporated into the plan.

In validating the manpower and material requirements, the reviewer is advised to look not only at global figures, but to go through unit calculations as well. For example, if the plan calls for 100 interviewers, he will wish to think about whether and how 100 suitable interviewers can be found, but also whether 100 is a reasonable number. The unit audit approach makes such judgments as what is the average time for a single interview, the travel time between interviews, proper hours of work per day per interviewer, days of work per week or month per interviewer, likely turnover rate of interviewers, i.e., how many interviewers must be replaced per month. He will wish to weigh these estimates against the total number of interviews to be made, and the time allowed for the task. The administrative aide will be concentrating on the supply side of the markets for space, equipment, and materials, and on procurement procedures.

The planned administrative, supervisory, and operational structure merits inspection from the point of view of sound and efficient management. Do the central organization, the regional field supervision, and the local operating arrangements fit the plan? Is the chain of command clear-cut? Is there a system of progress reporting? Are there pathways for resolving unforeseen problems in a manner that does minimum violence to the overall plan? Is the survey director in a position to exercise real administrative control over the entire operation?

The proposed timetable for conduct of the survey should be given a new scrutiny in the light of all other checking that has taken place. The examination will seek to confirm that it both meets requirements of the survey sponsors, and passes critical challenges as to feasibility.

Hopefully the overall review of availability of resources and of logistical arrangements finds the plan sound in these respects. It it does, fine. If it does not, then a further reconciliation is called for, this time to bring the newly validated design into agreement with a possibly modified set of objectives. Just as was true at the conclusion of the initial series of reviews discussed in section 5 above, any painful aspects of this reconciliation will be less severe

at this point in the survey than they would be if postponed to a later stage.

7. Control Over Data Collection

7.1 Planning and Operations. Once upon a time, a complaining cockroach who felt badly mistreated sought advice from a bedbug, who seemed to the cockroach to live a much more comfortable and pleasant life. After a little consideration, the bedbug responded, "Yes, I can help. My advice to you is to become a bedbug like me." The cockroach said, "Fine. How do I do it?" The bedbug replied, "That's your problem. That is operations. I'm concerned with planning!" Thus far in this manual we have discussed planning, and tried to point the way to validating a chosen plan. That is an important part of overall operational control of the survey, but as was noted earlier, it is in the production stages and especially in the actual collection of data that managerial control is tested most severely, for it is here that intentions are most likely to be violated, and so we turn to some of the steps which can be taken to reduce the risk that faulty collection of data will impair or destroy the survey. The essence of the approach now is not to try for perfection, but to attempt to insure that the data will be collected in reasonable conformity with the validated, chosen plan. We are no longer saying how the job *should* be done, but only trying to guarantee that it *is* done as planned.

Section 7 presents and discusses control mechanisms for two separate phases of data collection. Sections 7.2, 7.3, 7.4 and 7.5 are concerned with a group of control and quality-enhancing functions which are key elements of the data collecting process, but which precede the main body of field enumeration. Section 7.7 on sampling controls proposes devices of which some are pre-enumeration controls, and some are concurrent with field enumeration. The remainder of the chapter is devoted to controls over the major interviewing or enumerative operation itself.

7.2 Pre-Tests. Section 5.2 introduced in broad outline an illustration in which an household interview survey of about 10 per cent of the population of the province of West Dakota was to be conducted. The survey would determine population size, change in population, health status, and work status of the people. Since specific operational control procedures should be adapted to the particular survey, and in order to talk in concrete terms, we shall have a survey of that type in mind as we describe control processes which should be a part of data collection.

Planning will have produced a draft questionnaire and a preliminary set of instructions for interviewing, and for transcription and collection of responses. It is not unlikely that two or more different versions of the questionnaires will have been drafted. It is almost certain that a single first draft will not be a perfect instrument in every respect. Whether there is a single first draft, or several, there needs to be field trials to discover ambiguities, omissions, misunderstandings, irritations to respondents, awkward sequences of queries, and other weaknesses in the original drafts. Typically these pretests can and should be small-scale trials. Often they can be sequential in nature. The interviewing should be done by experienced personnel at this stage — perhaps supervisors, or enumerators who have worked earlier in similar surveys. The persons interviewed should be members of the same universe as the sample cases, but need not be chosen randomly. In fact, it will be better to select subjectively a small sample which will reflect what are guessed to be extremes in the population, in order to get a wider range of test situations.

A "small" sample is recommended. No universal specification can be given of just what size is best. This will vary with the complexity of the questionnaire, the variety of types of respondents, and the skill with which planning has been done. Something like the following may fit the situation.

A. At least three different interviewers should participate.

B. In a first step each interviewer does one day's interviewing: perhaps 8 to 12 households.

C. The interviewers assemble with supervisors and the survey managers to compare notes, discuss problems, decide on any changes that are indicated for the questions, the manner of their being asked, or recorded.

D. A second day's interviewing, similar to that of the first day takes place.

E. Step C is repeated for the second day of interviewing.

F. A third day of interviewing — and perhaps a fourth day — followed by further discussion and possibly further decisions or change may complete the pretesting and result in firming up a semi-final questionnaire and collection procedure.

NOTE: The survey manager may wish to do some of the pre-test interviewing himself in order to get a first-hand impression of the climate of the operation and leads on some of the possible problems. This is particularly likely in the smaller scale survey, where less elaborately organized pre-testing is warranted.

In the pretesting just outlined, the participants must be alert for a wide variety of possible difficulties and other factors which might endanger (or promote) success in collecting needed data. The following list is in no sense exhaustive, but should be suggestive of elements to be observed.

(1) Were there problems in finding the designated sample cases? Were the instructions adequate for identifying and delineating the sample household?

(2) Was the interaction between the interviewer and respondent satisfactory, or could something be done to improve the contact? Was the respondent puzzled, or frightened, or angry at being questioned? Did the respondent appear to cooperate and try to respond as requested?

(3) From the interviewer's point of view did the interview move smoothly? Were the questions easy to ask, did they have the right sequence? Was the interviewer embarrassed or uncertain how to proceed at any stage? Did he feel that he had succeeded in asking the questions in the manner which had been hoped for?

(4) Were there questions which the respondent seemed to fail to understand?

(5) Were there questions which the respondent failed to answer because he did not possess the necessary information?

(6) Were there questions which the interviewer suspected were answered incorrectly, or not at all, because the respondent did not wish to tell the truth?

(7) Were there situations encountered in which the question, even though answered correctly in the literal sense, failed to elicit the desired information? For example, the questionnaire might have asked, "Were you unable to work because of illness?" An answer of "yes" might have been given by a well person because he had to take care of children who were ill.

(8) Were all the time references properly interpreted? Were distinctions clear cut among such points and intervals as "now," "usually," "ever," "past week," "three months," "year ago," "1970"?

(9) Were there difficulties in establishing proper relationships between respondents and subjects? Do the instructions seem to be well chosen in specifying which sample persons should respond for themselves, and for which another person should respond?

(10) In replies to different questions was there any internal evidence of inconsistency or confirmation? (A well designed questionnaire often will provide opportunities for this type of check. For example, through comparison of age and year of birth, or in listing of household membership and noting births of babies, or in identifying accidents and loss of time from work.)

(11) Did time-of-day of the interview seem to have an impact on results? How did it affect willingness or ability to respond? Is some period of the day to be avoided or preferred?

(12) Was there any evidence of better or poorer response if other members of a household were present when questions were being asked of a particular sample person?

(13) How convenient or inconvenient were the physical features of the questionnaire and the forms for transcribing replies? Was spacing and question arrangement easy to follow and read? Was typesize appropriate for lighting conditions encountered? Was paper of good substance and handy size? Was space adequate for recording responses?

(14) For purposes of subsequent scheduling, a record of times required for interview, for travel to and between interviews, and for other

tasks should be maintained during the pretest. This will be of great value in determining final requirements for numbers of personnel, sizes of assignments, and reasonable quotas for assessing performance.

7.3 Public Relations. Certainly good work in the planning phase will have recognized the importance of establishing a favorable climate of attitudes in which the survey is to be conducted and will have made provisions for promoting that climate. Pretesting may have uncovered further facets of public relations that deserve attention. "Public Relations" as used here reaches over a broad spectrum. It covers the areas of legal or official authorization, relationships with associated or affiliated or possibly competitive organizations, respondents, prospective consumers of survey results, the communications media, and the general population. Any step that can be taken to secure support of the survey by an affected group, or to prevent or neutralize opposition increases the chances for a successful venture. The conduct of a survey that is popular is likely to be a pleasant operation. Execution in an unfriendly atmosphere is difficult and in some circumstances nearly impossible. To secure the favorable climate may well be worth substantial expenditure of even limited resources. Attention should be given not only to *what* is done in this area, but also to *when.*

7.4 Recruitment and Training of Agents. One of the truisms of production is "Any job is only as good as the people who do it." Planning and pretesting have determined how many of what kinds of personnel are needed. Personnel management then encompasses setting specifications for the various positions, recruiting, hiring, training, testing, final selection, and supervision.

We assume that adequate recruitment has been accomplished. Training will require orientation of collection agents — and especially of the field supervisors and interviewers — with respect to the purposes of the survey, and an understanding of the policies and perspectives of the survey managers. It is recommended that collection personnel be given some promotional materials, and at least one inspirational lecture by a prominent person, or in any event by a person who has the respect of the trainees, as a part of the effort to cause these people to "believe" in the survey, and to be motivated to do an excellent job. Experience has shown that the motivated worker who is trying his best will accomplish more than the person who may have more inherent ability, but who lacks interest or conviction in performing the task.

But spirit and good morale are not enough. The interviewer, coder, clerk, and supervisor must learn the specifics of their assignments, too. They cannot do a satisfactory job until they know what to do and how to do it. First, then, comes formal training in the content of the instruction manuals. This is followed by testing knowledge of that content, both through written exercises and examinations and through trial performance. The testing is geared to identifying satisfactory performance, need for further training, and dismissal of those employees who show insufficient promise.

The initial training may include lectures, demonstrations by supervisors or leader-agents, question-and-answer exercises (without grading at first, in order to minimize fear of failure, and to develop confidence as the trainee finds he *can* do his task). Small groups of trainees working by themselves, quizzing and criticizing one another can be helpful.

Performance testing of interviewers should be done under close supervision and observance by supervisors. Evaluation is facilitated through the use of a check-list observation report by the supervisors.

Performance testing of coders, transcribers, and other clerks is easier and can be accomplished with structural exercises and during processing. These exercises will have known correct answers, and thus permit the calculation of error rates that in turn are definitive measures of performance.

Amounts of training and testing will vary with the initial ability of trainees, complexity of the tasks, adequacy of training materials, and the tolerance requirements. As for other components of statistical operations, data collection and processing performance is subject to hazards of both variance and bias. While it will not be possible to perform a fully desirable

arithmetic analysis of the problem, it will be useful in setting standards to think in terms of mean square error (MSE).

The MSE of an estimate x' of a parameter x was defined earlier as

$$MSE_{x'} = \sigma_{x'}^2 + B_{x'}^2$$

where $\sigma_{x'}^2$ is the overall variance of x', and $B_{x'}^2$ is the square of the bias; bias being the difference between the expected value of x' and the defined target parameter x. Mean square error can be expressed in several alternative forms. One that is instructive, and convenient for the discussion that is to follow is:

$$MSE = \sigma_s^2 + \sigma_m^2 + B^2$$

where: σ_s^2 is "sampling" variance or the variability among possible sample estimates that arises because not every unit in the population is included in the sample,

σ_m^2 is "measurement" variance, or the variability in the enumeration or measuring process that would arise in repeated trials of the survey, even for identical samples of units, and

B is bias as defined above.

Where these terms are divided by the square of the expected value of the estimate, x', (or more loosely by x^2), the absolutes are converted to relatives. The usual terminology is

$$\frac{\sigma_s^2}{x^2} = V_s^2 = \text{sampling relative variance, or sampling relvariance}$$

$$\frac{\sigma_m^2}{x^2} = V_m^2 = \text{measurement relvariance, and}$$

$$\frac{B^2}{x^2} = \text{square of relative bias.}$$

The square roots of relvariances are called relative standard errors.

Under ideal circumstances good survey design will have obtained the best answer which circumstances permit, to a very difficult question: namely, the allocation of available resources to (1) sample structure and size, (2) methods for reducing measurement variability, and (3) methods of limiting bias, all in such a manner that mean square error is minimized. The solution will have included speculatively determined targets for permissable average bias in the interviewing process, and in variability of performance among acceptable interviewers. Pre-testing may have strengthened these speculative determinations. The task in selecting and training interviewers is to meet these specifications as closely as possible for a few of the most important items in the survey. Under these ideal circumstances, that means training until the *average* performance of those interviewers who are to be retained differs from perfection by not more than the specified bias; and it means dismissing from employment or retraining those interviewers whose performance differs from the average by more than about three standard deviations when all candidates for the job are arrayed in a frequency distribution according to their respective performance scores. This ideal state of affairs is rarely if ever fully satisfied. In practice, an approach which is guided by the model just described can do much to assure that a reasonably adequate set of interviewers is chosen to initiate data collection. (See Appendix B for a suggestive illustration of how one may proceed.) In Section 7 below, procedures are recommended for maintaining control over subsequent performance of interviewers.

7.5 Dress Rehearsal. The survey has now reached the stage at which it is nearly ready "to go." Planning of operations has been completed, instructions written, personnel hired and trained, other resources and materials provided, pretesting of questionnaire and interviewing accomplished. Plans for editing, coding, data processing, and tabulation also have been set, although later sections of this manual will treat control of data reduction in more detail. A final dress rehearsal operation is advisable before entering into formal full production of the official set of survey data.

The dress rehearsal should be a close approach to a trial or experimental conduct of the final survey itself, except carried out in miniature and over a limited period of time. The objectives include again those discussed above

in section 7.2 dealing with pre-testing, although in this step it is hoped that already proven procedures are being tested, so that few further corrections are necessary. In the dress rehearsal there are the additional objectives of testing and revising as indicated the planned administrative and logistical features of the project, and of preliminary validating of data processing procedures.

The appropriate scale, extent, and detailed content of the dress rehearsal will vary with these same factors in the main survey. The greater the main survey, the more comprehensive the dress rehearsal should be. Remember that the central concept is to approximate the operational conduct of the main survey on a smaller scale to discover if the plan is feasible, to remove "bugs" which are found in actual trial, and thus to embark finally with machinery which has been "road tested." This is the same concept which has proved successful in manufacturing industries around the world.

When the main survey is a major undertaking, the ideal course is to run the dress rehearsal through one full trial time period or cycle of the pattern of the survey itself. For example, if the main survey is to be a continuing operation, or if it is to extend over a year or more of data collection, it likely will have been designed with subsamples of places or of months (or quarters) of data collection, or of both places and time. The dress rehearsal might then cover all — or perhaps a random fraction — of one corresponding subsample, or one month of trial collection. Data from the dress rehearsal will be collected, transcribed, edited and coded, and possibly tabulated as planned for the main survey, but they will *not* be incorporated into final survey results, except perhaps as preliminary and provisional data. This is the last opportunity to correct procedural instructions for collecting and processing the data; to revise editing, coding and classification structures; to introduce devices to compensate for earlier omissions in instructions.

Managers will be observing the dress rehearsal on the administrative and logistical side for feasibility and for adequacy. Particular attention will be given to time tables, costs, transportation and communication problems, transmittal and transcription of data, flow of supplies, personnel management, and channels of supervision. Each of these elements can be critical, but perhaps the most important of all are the latter two. Lines of authority and supervision must be both clear-cut and firm. A disciplined but happy crew of interviewers and other agents is an absolute essential to success. Every interviewer should feel this discipline, but he should also believe that the system is not oblivious to his welfare, and does understand his problems. He should know where and how to turn for help when help is needed.

7.6 Systems of Controls over Field Collection. The main survey is underway. Planners have done all they can to prepare for a sound undertaking. How can it be kept on the track? Desire, motivation, conviction, dedication, perseverance, courage, and careful attention to task on the part of all personnel are needed. The extent to which they are present will largely determine the answer. Fortunately, there are a number of tools which can facilitate the job. Again we note that the choice of tool will vary with type of survey, and particularly with the mechanism for primary collection of data, e.g., whether it be interview of persons, transcription of records, observation, or new physical measurement. In this section, we concentrate on the interview instrument, and rely heavily on the power of replication as a control device. The essence of recommended control procedures can be adapted to other data collection procedures — for instance, transcription of records — without great difficulty, and indeed often without any substantial change.

For operations of the general type of the West Dakota survey previously discussed, recommended controls over data collection fall into six classes.

A. Control over sample inclusion, exercised through re-selection of subsamples.

B. Control of the interviewing process and of individual interviewers through reinterview of a subsample of respondents.

C. Control of the interviewing process and of individual interviewers through observation by supervisors of a sample of interviewer assignments.

D. Control of the interviewing process and of individual interviewers through edit checks for performance.

E. Minimization of nonresponse through a system of surveillance of nonresponse rates, assigned quotas, and, if necessary, introduction of additional follow-up procedures.

F. Screening early survey results for internal consistency and correspondence with any available external data, and intensifying efforts in any area of apparent discrepancy.

Classes A and E should be a part of all surveys. The other classes of control may be used individually to greater or lesser extent, or in combination. But whatever specific overall methods are adopted, they should include control over the four factors of sampling, nonresponse, the interviewing *process,* and performance of individual interviewers.

7.7 Sampling Controls. Validation of sampling operations has two phases. The first is verification of the mechanics of selection of first-stage primary sampling units (PSU), and of any intermediate sampling units for which sampling is carried out in central offices. The second relates to ultimate stages of sampling performed on the field by enumerators or interviewers. We treat central office sampling controls first.

This control is essentially a verification of the mechanics of selection. It can be accomplished through either of two approaches or variants of those approaches. In one scheme, the work of sample selection of PSU's (and intermediate units, if they are a part of the design) is redone independently by a second clerk or group of clerks. Any discrepancies between the two selections are reconciled by a supervisor, or if necessary, by the managing statistician. Possible difficulties may involve only clerical errors or such things as cumulative impact of rounding errors, or misinterpretation of an ambiguous instruction. But they may have more hard-to-resolve origins. For example, if the PSU's are defined in terms of geographic and political boundaries, it is not unlikely that there will be instances in which proper boundaries are uncertain. These situations must be resolved in a manner that does not bias the sampling. If any part of the sampling is from listed frames, care must be exercised that the target universe is exhausted by the lists used, or not partially duplicated. It is easier than one might at first suppose for clerks to miss one of 24 file drawers in drawing sample cases, and especially if one or more of the drawers is a "supplement" to the main file, a group of cases "inactive" or "in process," or in some other respect different from the central frame. When there are undetected duplications in the frame, the realized probabilities are different from those intended, and subsequent messy field episodes and computational adjustments must be faced.

A second approach may be used alternatively, or in conjunction with the dependent verification just described. It is easy to apply in some situations, and much more difficult, or perhaps impossible, in others. In this approach, one makes calculations to see if the realized sample does indeed yield a statistic which is a satisfactory estimate of a criterion parameter of the target universe. The procedure is outlined with the following illustration. Specific details of application will differ with particular designs, and depend upon the nature of available data. The chosen illustration is a simple one, in order to more clearly display the highlights of the method.

A. Assume that the universe consists of N PSU's with z_i being the criterion measure for the i^{th} PSU. The z_i statistic may be — and preferably is — a known value from some previous statistical or administrative program. For instance, it might be the head count in an earlier population census; or the numbers in a registry; or tax records; or even estimated or guessed numbers of children born. The z_i measures must be in specific numbers, but it is not necessary that they be accurate measures. If a choice is possible, it is better to choose a z_i-measure which can be expected to be positively correlated with x_i-values, where x_i is an (unknown) key statistic and objective of the survey. If no z_i real data can be found, it is still acceptable to assign arbitrary z_i values to the PSU's, choosing z_i numbers which one suspects are correlated with the unknown x_i's. This assignment of synthetic data to the PSU's can be done with considerable sophistication, using the same

type of evidence that was given weight in designing the sampling plan, taking into account expected mean values and speculated population variances, with specific selected z_i's being chosen randomly from appropriate distributions of z_i-values. A less elaborate procedure which bears a resemblance to the outlined one may well be adequate. The assignments must, of course, be made by someone who does not have knowledge of the specific PSU's which have been selected in the sample.

B. For the *n* PSU's in the sample, transcribe the respective z_i-values that have been chosen in step A. The assumption implicit in this new step is that subsequent stages of sampling within the selected PSU's either would yield precise measures of z_i for the PSU, or, since we are attempting at this point to verify only the first stages of sampling, the subsequent sampling is not of immediate concern.

C. Using the z_i-data for the *n* sample PSU's, and the full intended estimating procedure of the survey design (minus the within-PSU elements), calculate z', an estimate of z, the aggregate total value of the criterion statistic for the full target universe.

D. The discrepancy, $d = z' - z$, should come from the normal distribution with mean zero, and variance $\sigma_{z'}^2$, which can be calculated from the assigned values z_i in the full universe, or estimated from the assigned z_i-values in the sample.

E. If d is less than $\sigma_{z'}$, sampling in the first stage can be considered verified. If $2\sigma_{z'} \leqq d \leqq 2\sigma_{z'}$, the sampling may be correct, but merits some further spot checking. If $|d| > 3\sigma_{z'}$, probably errors have been made in sampling, and a very careful review is warranted.

The second phase of sampling control is concerned with selection of ultimate sampling units in the field by enumerators or interviewers. Control of this process is best handled in combination with the Class B control system, i.e., through re-sampling and reinterview of a subsample of assignments. For convenience in discussion, the entire process is designated the control subsample and also will be referred to at times simply as the reinterview.

7.8 The Control Subsample. The primary purposes of the Control Subsample are the three already identified:

A. Confirmation of field sampling and proper inclusion of ultimate units in the sample,

B. Surveillance of the interviewing process (both as a mechanism for formal process control and as a motivating force for interviewers to perform in a manner which will pass inspection), and

C. Control over individual interviewers, identifying those who are performing satisfactorily, those who need retraining on selected matters, and those who must be replaced.

In addition, the reinterview may serve two other purposes:

D. Strengthening the capability of field supervisors, and

E. Providing evidence which later will assist analysts in interpreting survey results, and in estimating overall mean square errors.

It should be said immediately that operation of a control subsample is itself a difficult matter. It requires honest and diligent effort by field supervisors. It calls for careful and tight scheduling. It will lose much of its value if the reinterview and processing of evidence from the reinterview do not follow closely in time the original interviews. There must be prompt feedbacks from reinterview findings to production operations if the control is to have full effect.

The general plan of the control sample is simple and straightforward. It consists of (1) drawing a subsample of assignments to interviewers; (2) re-doing by experts the work that had been intended for the subsample of assignments; (3) comparing the results from the original assignments with those from the check assignments for sample coverage, data content of key items, and for variation among interviewers in performance; and (4) taking any indicated action to bring the main survey into closer conformance with the survey design. While the strategy is simple, the tactics or details of operation of the control subsample are not amenable to simple or unique specifications. An illustrative arrangement is offered below, with attention being called to several particularly significant aspects.

Suppose field work is built around the following elements. (The specific numbers mentioned for each parameter are included in order to make parts of the illustration more concrete. They are not necessarily recommended quantities): The discussion makes use of an important design and operational element called an "assignment." An assignment is a specific piece of work given to an interviewer (or reinterviewer) to be accomplished in a stated period of time — often a day or a week in survey programs. Not uncommonly, it will consist of collecting data from a cluster of 10 to 25 sampling units in geographic proximity to one another.

M = number of interviewer assignments for one time period or cycle of the main survey — perhaps a 3-month period — perhaps 500 assignments.

H = number of households in the main sample in this period — perhaps 10,000.

N = number of persons in the main sample in this period — say 40,000.

$\bar{H} = H/M$ = average number of households per assignment in main sample — 20.

$\bar{N} = N/M$ = average number of sample persons per assignment — 80.

K = number of interviewers — say 50.

$\bar{M} = M/K$ = average number of assignments per interviewer in the time period — this can vary widely from one survey to another; for the illustration it is 10.

R = number of expert reinterviewers—say 10.

m = number of interviewer assignments in the control subsample.

h = number of households in the control or reinterview subsample.

n = number of persons reinterviewed.

$\bar{h} = h/m$ = number of households per assignment in the reinterview.

$\bar{n} = n/m$ = average number of reinterviewed persons per assignment.

$\bar{\bar{n}} = n/h$ = number of persons per household reinterviewed.

$\bar{m} = m/K$ = average number of reinterviewed assignments per interviewer.

It is assumed that work loads are such that all expert reinterviewers — probably the field supervisors — participate equally in the reinterview procedures, each handling $K/R = 5$ interviewers.

A number of decisions must be made; most of them being subjective judgments since firm planning data which would permit more rigorous determinations will be lacking. Clearly one decision must be to make m, h, and n large enough to provide sufficient data to be useful for any purpose. The sizes of m, h, and n are limited by several factors, and especially by cost and by the demands on the time of supervisors. It is certainly to be hoped that most original interviewers will miss not more than a small proportion of persons who should be in the sample. This means that for the coverage check, a moderately large subsample is needed to detect a significant omission rate.

A start toward determination of subsample "size" might be made in this way: Perhaps a tolerable omission rate is 5%. Managers wish to make reasonably sure that if the actual omission rate is double that, it is not measured as low as 2%, and thus incorrectly accepted. Translate "reasonably sure" to mean there should be only a 5% chance that an effective probability of omission 0.04 is measured to be as low as 0.02. The control procedure is concerned with overcoverage as well as undercoverage, but since undercoverage is the more serious problem, in planning one is reasonably sure that if he provides adequately for detecting omissions, he will do at least passably well on the other side. While some persons may be missed within a household, the important risk of omission is of entire households, and so subsample size is approached initially in terms of households. Assuming as an adequate model simple random sampling, relevant calculations are (with P = *probability of omission*, $Q = 1 - P$, σ^2 = *variance of estimated P*, h = *number of households*):

$$1.65\sigma = 0.04 - 0.02$$

$$\sigma \approx 0.012$$

$$h = \frac{PQ}{\sigma^2} \approx \frac{0.04(0.96)}{0.000144} \approx 267.$$

This suggests that it will require a subsample of approximately 260 households or 13 full assignments to meet the suggested coverage

check. It is immediately clear that the control subsample cannot provide in one cycle an adequate coverage check for each interviewer, for that would require more than a 100% reinterview. But if reinterviewing is spread out evenly among supervisors, interviewers and time, one reinterviewed assignment per interviewer would easily meet the overall coverage check each month. This suggests a tentative plan for recheck of five assignments by each supervisor each cycle, covering one assignment of each of the interviewers under his supervision every three months. This is a reasonable work-load for the supervisor, and it has the attractive administrative simplicity of one check per interviewer each cycle. The latter point exhibits one of the intrinsic strengths of the reinterview program. Assuming that the choice of recheck assignment for an interviewer is made randomly from among each 10 of his assignments, the interviewer is under continuous risk of having his work checked, and thus is motivated at all times to do a better job. The incentive is especially obvious if the interviewer's compensation is influenced at all by the quality of his performance.

How does this tentative scheme fit other objectives? Some further procedural questions must be answered first. The full assignment was checked for coverage because it is proper inclusion of households that is the most critical element of the check. For reasons of minimizing travel cost, and completing the coverage check of persons, it is desirable to also reinterview all households in the recheck assignments for sample membership. The situation with respect to detailed reinterview of persons for subject content is different. Here several new questions arise. In the original interview one respondent may have replied for all members of the household. There is likely to be an intra-household correlation in accuracy of reporting for different household members, and some of the data applies to the household as a whole rather than to individual persons in any event. It appears advisable to conduct the reinterview with respect to only the household questions and one subject person per household.

Thus the tentative plan embraces interview of only 1000 persons per cycle, an average of 20 respondents per interviewer, and 100 respondents per supervisor. Completed interviews will be a bit less than these figures because of unavoidable non-response.

A number of variations and modifications of the above plan will occur to the statistician faced with a particular set of conditions. In evaluating a chosen scheme, it is especially important to realize that the scale of activity in the control subsample must for reasons of cost and feasibility be modest. There will at few points be a sufficient number of observations to be absolutely sure of conclusions. Accordingly, the reinterview procedure must be considered as only a part of the overall control of field operations. It can do a great deal, but not the whole job.

This manual is not a textbook on formal process or quality control procedures as they have been developed in industrial production. The student who is interested in optimal designs and in more elaborate interpretation of the evidence coming from those designs is referred to one of the several good textbooks — for example, references [4] or [7]. Some additional comments on formal quality control are offered in section 8 below, "Data Processing." A few notes on formal quality control are presented in Appendix A. A thorough grasp of formal quality control theory would permit a survey manager to introduce some refinements and efficiencies into control of the interviewing process and of interviewers, but is by no means necessary. A more pedestrian approach is taken in here. This course is justified by three main considerations. First, the body of theory which underlies formal quality control designs is extensive, highly mathematical, and complex. It is a field in itself. Thorough understanding requires either very considerable study or already trained personnel who may not be available. Second, the nature of the interviewing process is such that exact precision of measurement and determination of parameters needed for most formal quality control models is not obtainable. A greater element of subjectivity enters the picture. Third, and perhaps most important, the design of an effective reinterview program is determined better by careful attention to administrative factors and opera-

tional feasibility, than by too slavish adherence to a plan that promises but still may not be able to deliver some specified numerical level of accuracy.

Indeed, the indirect effects of quality control procedures such as reinterviewing and replicate coding are likely to be more important than the direct effects. For as noted earlier, the knowledge by interviewers and coders that their work is to be checked is a powerful incentive to do the job correctly in the first place — and certainly to avoid deliberate faking or "curbstoning," a term sometimes used to refer to the dishonest action of an interviewer who sits at home or on a curbstone and fills out a questionnaire without actually talking to the named respondent.

We proceed to comment further on how evidence from the control subsample described in this section can be utilized.

For operational purposes, a simple model provides an adequate key instrument for evaluating performance, and suggesting appropriate action — whether the focus of attention is on a single interviewer, the group of interviewers under a common supervisor, or on a particular question or set of questions in the survey. Let n_c be the number of observations of the item or category c under study; d_c be the number of errors observed for that item (i.e., the number of instances in which the reinterviewer's finding differs from that of the interviewer). Then $P_c' = d_c/n_c$ may be treated as an estimate of the probability of error for that item for the category of n_c observations being considered. The estimate P'_c can be assumed to have a simple binomial distribution with variance $\sigma^2_{P'_c} = P_c' (1 - P_c')/n_c$. Consider, now, some of the uses of this tool.

It will be convenient to assemble data from the control subsample in tables which have a standard format, similar to that illustrated in Table 7-1. Table 7-1 is a generalized table, illustrated for evaluation of the reporting of numbers of persons with one or more chronic conditions. It can be easily adapted to the coverage check, or to proper reporting on any specified topic, group of questions, or classes or types of respondents or subject persons. Tables like 7-1 can be consolidated, and summed over interviewers, reinterviewers, question categories, or time intervals.

The symbols appearing in the column heads correspond to those in the simple model just described. Note that the table allows for two counts by the reinterviewer, y_c' and y_c, the first his initial determination, and the second his final determination after "reconciliation." While there are acceptable variations, the table presumes a procedure in which the reinterviewer first records his finding without knowing what the original interviewer reported; then consults the report of the interviewer; compares the two, and reaches a final conclusion, y_c, perhaps consulting further with the respondent, and perhaps amending his own initial finding. This procedure is recommended for these reasons: (1) Although fully independent verification procedures are preferred for most operations, in the reinterview situation, the presence of the reinterviewer at the location of the respondent at the time of reinterview offers an opportunity for reducing misunderstandings through the reconciliation process. (2) The reinterviewer, although a presumed "expert" may be little if any more skillful than the better interviewer, and may properly be influenced in making his final judgment by taking account of the opinion of the interviewer. (This does put some strain on the integrity of the reinterviewer who is the interviewer's supervisor, but higher level supervision is likely to detect misbehavior in this respect should observed differences over an extended time range be too much smaller than expected.) (3) The reconciliation steps are a valuable training device for the *re*-interviewer, in making him more intimately aware of the difficulties which interviewers are having.

The table includes both the algebraic difference, d_c, and the absolute value $|d_c|$. This permits summation to reflect either gross or net differences, and, of course, points the direction in which re-training or other corrective actions should move.

As noted earlier, when the focus of attention is on an individual interviewer, a single cycle of reinterview (in the illustration), offers only crude evidence of performance. When the statistic of interest is a person characteristic, the

single cycle provides only 20 observations. An estimated error rate of 30 per cent, for example, has a sampling error of 10 percentage points. A 30 per cent error rate is intolerable, and 10 percentage points in sampling error leaves too much doubt. But consolidation of evidence can improve the picture markedly. The interviewer who makes anything like 30 per cent error on one item probably has errors of similar scale on, say, 10 other items. In that event, the sampling error for the group of items is cut to about 3 percentage points, and it becomes pretty clear that the interviewer is performing in an inadequate manner. Combining observations for the interviewer over several cycles of reinterview has a similar utility. Comparison of the estimated error rates among interviewers may well isolate the poor performer. Use of this rather subjective pattern of evaluation in conjunction with the simple binomial model described above, is the approach recommended in this manual. When an interviewer shows unsatisfactory performance, he must be retrained — and that is preferable, if feasible — or replaced.

Some may prefer more structured guidance for interpretation of reinterview data. Such is offered below, with the caution that it still has arbitrary and subjective features. It is a product of standards which could be set at different levels and is an adaptation of an "acceptance sampling plan."

Step One. Identify Sections I, II, III, . . . of the questionnaire, for which separate surveil-

Table 7-1. Re-Interview Data on Number of Persons with One or More Chronic Health Conditions for District A, West Dakota Survey, Cycle 7

Stub	For Subject Persons Reinterviewed					
	Number Persons	Number Persons with one or more chronic conditions				
		Reported by Interviewer	Reported by Reinterview		Difference	
			Initially	After Reconciliation		
c	n_c	x_c	y_c'	y_c	x_c-y_c	$\lvert x_c-y_c \rvert$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interviewer 1 Interviewer 2 Interviewer 3 Interviewer 4 Interviewer 5						
or Question 1 Question 2 Question 3						
or Respondent Type 1 Respondent Type 2 Respondent Type 3						

lance is desired. For example, some of the sections might be:

Section I Coverage of households.
Section II Coverage of persons.
Section III Questions on demographic characteristics of individual persons.
Section IV Health or Labor Force attributes of persons.

Step Two. For each section, identify and count the units of observation, i.e., the effective sample size — usually number of households, number of persons, or number of person-questions. (This is n_c for a category.)

Step Three. For each section select an approximate level of acceptable error rate, i.e., a target value of P_c.

Step Four. Define as *Acceptance Number* the maximum number of permissible differences, d_c, for a given range of observations, n_c, for each section.

Step Five. Build a table like Table 7.2, "Acceptable Difference Limits" for each section.

Step Six. A decision is made to "accept" or "reject" the work of each interviewer for each reinterviewed assignment. If the work is accepted, no further action is taken. If rejected, the interviewer must be retrained on the rejected section(s), or replaced.

Table 7.2 was built around a target acceptable quality level (AQL) of 5 per cent error rate. This means that if the true average error rate is 5 percent or less, the performance will be accepted at least 95 percent of the time. The figures in the right-hand column are approximations to what is sometimes called "the consumer's risk." They are interpreted in the following way: If, for example, there are 60 observations, and the acceptance number 6 is used, an actual underlying error rate of 18 per cent or greater will not be accepted on more than 10 per cent of inspections. It may appear that the acceptance numbers in the table seem to reflect error tolerances which are nearly double the target specifications. The explanation lies in the fact that the acceptance numbers are *maximum* values. Observations with fewer errors will also be accepted, and so the average error rate will be smaller.

Table 7.2 Acceptable Difference Limits for Section IV

(For example, number of persons with specified health attributes)

Number of Observations n_c	Acceptable Number of Differences d_c	Approximate $P_{0.10}(100)$
10- 29	2	26%
30- 49	4	20%
50- 69	6	18%
70- 99	7	14%
100-129	9	13%

Those who might wish to set different standards in constructing a table similar to Table 7.2, may consult quality control textbooks. (cf. p. 358ff of reference [7] or p. 154 of reference [4].

An analogous procedure to that outlined for evaluating interviewers can be employed to assess whether a particular question or respondent type, or a group of questions or respondent types is being handled in satisfactory fashion.

It will be desirable for the purpose of surveillance of the overall interviewing process to construct and maintain a simple control chart. The chart gives a picture of the trend of a summary error rate over time and tells the survey manager whether the process is "in control."

Again there are alternative routes for establishing and using the control chart. The following scheme is straightforward, and provides an informative tool.

Step 1. Form a weighted average of the chosen target error rates over all sections of the questionnaire and all interviewers. Average the error rates among interviewers within questionnaire sections for a single cycle of the survey, and then form the weighted average of these section averages to yield

$$\overline{R} = \sum_{i=1}^{L} w_i R_i$$

Where R_i is the average target for all interviewers for section *i* of the questionnaire and interview process;
w_i is the assessed relative importance of section *i* in the overall process. For example, w_i may be assigned the value 0.30 where *i* is the cov-

erage process; w_i assigned 0.10 where *i* is the demographic attribute section; w_i assigned 0.20 for the health attribute section; . . . $\sum_i w_i = 1.00$;

L is the number of sections; and

$\overline{R}$ is the overall target error rate.

Step 2. For each cycle of the survey, compute the realized overall error rate $\overline{R'}$ with the formula

$$\overline{R'} = \sum_{i=1}^{L} w_i R_i'$$

where R_i' is the observed average error rate for the section, for a given cycle.

Step 3. Assuming that the interviewers have approximately equal work loads, an adequate estimate of the sampling variance of an observed error rate R_i' is

$$S^2_{R'_i} = \frac{R_i'(1 - R_i')}{n_i}$$

where n_i is the total number of observations over all interviewers for section *i* for one cycle. For example in the illustration on p. 25 for household coverage, n_i will be approximately 1000.

The variance of the estimate of realized overall rate, $\overline{R'}$ will be approximately $S^2_{\overline{R'}} = \sum_i w_i S^2_{R'_i}$.

The standard error of $\overline{R'}$ is the square root $S_{\overline{R'}}$.

Step 4. Set up the control chart as in Figure 7.1, and post the values of $\overline{R'}$ to the chart each cycle of the survey.

The center horizontal heavy line is the overall target acceptable error rate. The horizontal lines at $\overline{R} \pm 2S_{\overline{R'}}$ are the boundaries of tolerable observed error rates. As long as the plotted values $\overline{R'}$, fall between these boundaries, the interviewing process can be considered "in control." In the illustration in Figure 7.1, the steady climb in $\overline{R'}$ from cycle 6 through cycle 9 raises a suspicion of emerging trouble, but the process has still remained in control through cycle 9. At cycle 10, however, it moves out of control. At that point, prompt and firm action must be taken. The components of the weighted average $\overline{R'}$ which are principally responsible for the high error rate must be located, and corrective measures taken to bring the process back into control.

Evidence from the control subsample can serve one further important function. As the analyst studies final tabulations from the overall survey, attempts to draw conclusions and write reports for presentation to officials or for publication, his work will have greater validity if he is able to make sound judgments concerning the general reliability of specific statistics. Knowledge of sampling errors is a great help, and it is presumed he has that information. The reinterview data can give him another powerful tool, for they provide an indicator of the stability of the measurement process. This tool is not a precision instrument, but it is sufficiently sharp to permit the analyst to distinguish the consistently measured statistic from the one with highly unstable response — and consequently one with questionable validity. Data from a table similar to Table 7-1 provide the needed input. For a particular statistic — say percentage of persons who have had a specified disease — one can calculate for each interviewer for each cycle the rate $r' = x/n$ as estimated from interviewer reports, and the rate $r'' = y/n$ as estimated from reinterview reports. The difference $D = r' - r''$ is a first crude measure of consistency or stability in measurement of this item. The average squared value of this difference taken over all interviewers and all cycles,

$$S^2 = \frac{1}{K} \sum_{k=1}^{K} D^2_k$$

where the subscript *k* refers to observation for one interviewer in one cycle, and K is the total number of interviewer-cycles,

is a summary measure of the stability of the statistic. Statisticians will differ on the exact interpretation of the statistic S^2, but all will agree that it tends to rank order items with respect to the consistency of measurement. If *r* is the overall survey estimate of a particular rate, then $v = S/r$ is a relative measure of the inconsistency of measurement. One group of statisticians at the U. S. Bureau of the Census has developed a more elaborate model for

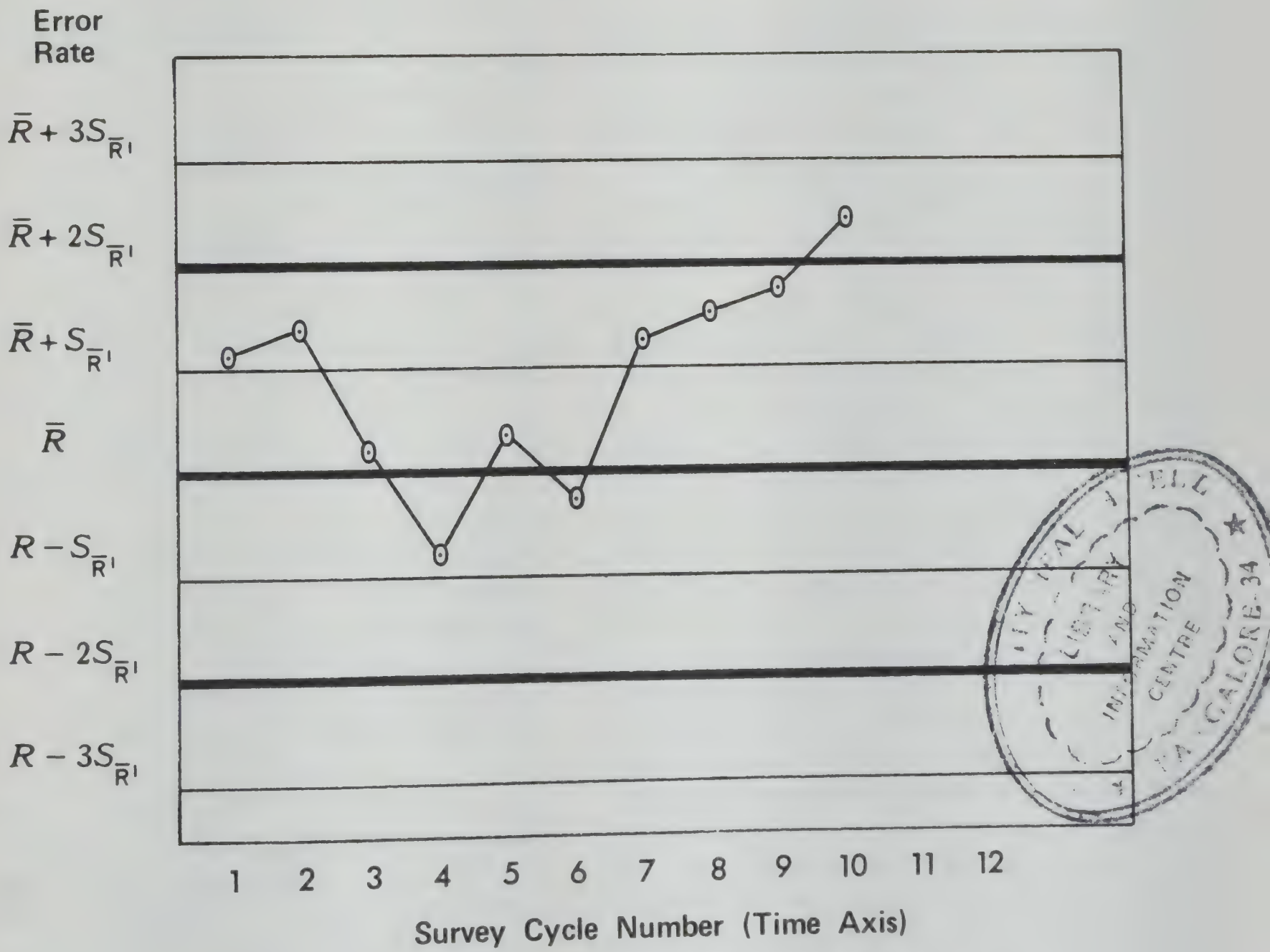
analyzing reinterview data for consistency or stability of particular statistics. See reference [6]. Their model begins with data being classified as in Table 7-3. In this table, data for each individual sample observation are recorded; the value 1 is assigned to an individual classified as having the specified characteristic, and the value 0 otherwise.

Thus of the n individuals reinterviewed, a had the specified characteristic in both interview and reinterview, d in neither, b in reinterview but not in interview, and c in interview but not in reinterview. The quantity $(b + c)$ *is* called *gross difference,* and $(a+ b) - (a + c) = (b - c)$ is called *net difference.* Gross and net rates are $(b + c)/n$ and $(b - c)/n$, respectively. Since either $(0 - 1)^2$ or $(1 - 0)^2$ is equal to unity, the mean sum of squared individual differences is simply the gross difference rate $g = (b + c)/n$. (This is a measure similar in concept to the S^2 of the previous section. It can be shown that the expected value of g is equal to twice the simple response variance. See reference [6].)

Table 7-3 Numbers of Individuals with and without Specified Characteristics on Interview and Reinterview for Identical Individuals

Results of Reinterview Survey	Results of Initial Survey		
	1	0	Totals
1	a	b	a+b
0	c	d	c+d
Totals	a+c	b+d	a+b+c+d=n

Figure 7.1. Control Chart, Interviewing Process



Each plotted point 0 is the observed overall error rate, $\bar{R}'$ for one cycle.

DEM-130 N71

One estimate of the overall population variance of the proportion of individuals with the specified characteristic is

$$S_1^2 = \left(\frac{a+b}{n}\right)\left(1 - \frac{a+b}{n}\right);$$

another is

$$S_2^2 = \left(\frac{a+c}{n}\right)\left(1 - \frac{a+c}{n}\right);$$

and a third in their average

$$S_3^2 = \tfrac{1}{2}\,(S_1^2 + S_2^2).$$

The model defines the ratio of *1/2* the gross difference rate g, to the estimate S_3^2 as the *Index of Inconsistency,*

$$I = g/2S_3^2,$$

and interprets this index, I, roughly as the proportion of total observed variability in the population which is due to instability of measurements.

7.9. An Observation Program. Some survey managers recommend a more highly subjective form of field control than the control subsample. They advise an interviewer observation program. This manual on operational control prefers to view an observation program as a supplement to rather than a substitute for the control subsample.

An observation program is a systematic scheduling of occasions on which a supervisor or an experienced senior agent accompanies an interviewer on an assignment, in order to secure through direct inspection more insightful knowledge of whether the interviewing process is being carried out as planned, and in order to be better prepared to take any corrective actions which may be indicated. The program has both training and control functions. It focuses attention on much the same matters as a reinterview system: completeness of coverage, proper asking of questions and recording of answers, performance of individual interviewers, apparent difficulties with certain questions or types of respondents. In addition, it gives the supervisor another opportunity to assess the environment in which the interviewing process takes place, and occasions for more intimate contact with interviewers, including discovery and discussion of points which other more structured techniques might overlook.

The scheduling of observations may take account of three situations. The first is an observation of a new interviewer shortly following completion of her initial training. The second is a periodic regular observation of one assignment or a part of an assignment at perhaps six-month intervals. (More frequently, if there is no reinterview system.) The third is special occasions when other evidence suggests an emerging problem, or the need for retraining.

The primary strength of an observation program should lie in its training aspect, and in its positive force to help the interviewer to do a better job. The interviewer is likely to be under some stress when being observed by a supervisor. The supervisor should be considerate — he should fill the role of teacher rather than policeman. Yet the process does have control features, for it motivates the interviewer to do his best, it guides the supervisor and manager toward improved procedures, and it may be a timely locator of substandard or worse performance by particular interviewers.

Specific scheduling of observations, with respect to point in the survey cycle, day or days of the week, time of day, or frequency, will depend not only on the general time table for survey operations and other demands on the supervisor, but also on assessment of what shortcomings are most likely to be found, i.e., whether the problem is low production rate, high non-response rate, higher error rates on particular items, or general inadequacy.

It is suggested that a supervisor's report be prepared for each observation, following a structured format. This puts a desirable element of discipline into the program, makes it a little less subjective, and provides higher level supervisors with materials which can be at least partially summarized. The report may be most easily executed on a standard form, but whether it is or not, it should contain items similar to those in the following list. Except in unusual circumstances, and except for a possible confidential portion, copies of the report should go to the interviewer, and to the central office.

(1) Interviewer name and identification number.

(2) Observor name and identification number.
(3) Date(s) of observation.
(4) Number of assignment observed.
(5) Type of observation
   A. Initial
   B. Regular
   C. Special
(6) Coverage of households probably accurate: Yes____; No____.
(7) Coverage of persons probably accurate: Yes____; No____.
(8) Non-Response:____ households from____ in sample.
(9) General Performance of Interviewer:

Facet	Good	Acceptable	Unsatisfactory
A. Personal Appearance and Behavior			
B. Securing Respondent Cooperation			
C. Following Established Procedures			

(10) Handling of Sections of Questionnaire:

Section	Good	Acceptable	Unsatisfactory
A. Household and Person Identification			
B. Demographic Descriptions			
C. Vital Events			
D. Labor Force Items			
E. Health Items			

The effectiveness of an observation program comes in considerable part from its indirect impact on both interviewer and supervisor. But it is enhanced by prompt feedback of findings through the survey machinery in order to bring about improvements where weaknesses are discovered.

7.10 Error Reports from Central Editing. Reinterviewing and observation are strong control mechanisms which utilize inspection and evaluation of samples of interviewer performance. One other surveillance procedure, which renders a more intensive check of certain aspects of reported data. is recommended. Section 8 of this manual will deal with editing and processing controls. It will be found that the recommended editing procedures yield counts of errors discovered by editors for each of a number of questionnaire items, when the nature of the item is such that a clerical check (by either a clerk or a computer routine) can identify an imperfection. These error counts should be tabulated by interviewer, and converted to summary error rates, for classes of questions, for each interviewer. Further summary by classes of questions across interviewers, and by supervision district, provide another assessment of conformity to specified rules by interviewers, supervisors, and for classes of questions.

These error summaries, like other control evidence, should be distributed promptly to appropriate points in the survey administration where they become the basis for corrective action. When subsequent reports do not show trend improvement in excessive error rates, disciplinary steps are in order, unless the error is so pervasive that a revised standard must be adopted.

7.11 Combining Control Evidence. Throughout section 7 a variety of methods have been described for exercising control over the collection of data — with special emphasis on both the overall process and on individual interviewer performance. It is quite possible one of these methods might suggest that a particular interviewer was unsatisfactory, and another classify him as acceptable. It would make for tidiness to propose some rigid formula for combining results from the several methods, which would tell the survey manager exactly and definitely what steps he should take. While that might be done, it would be unwise or even foolish. Rather, the manager and supervisors must study all the evidence which the various control mechanisms provide. When any one control device produces a substantial danger signal, the trouble should be investigated, for that is the purpose of the system. But in any given situation, there may be extenuating circumstances or misleading evidence. An interviewer may have been temporarily ill; a group of respondents abnormally uncooperative;

weather too severe. There might have been local civil disorder; some other unusual event might have produced a temporary disarrangement of the public temperament; the control report itself could have been in error. Guided by the evidence of the mechanisms, and using all collateral information at their command, they will be in a position to make sound judgments which should keep data collection under control. Generally speaking, it is better to retrain an interviewer than to dismiss him — thus saving the cost of recruiting and completely training a new person who may be as poor. It will be better to amend or reinforce instructions than to adopt some completely new course in mid-stream — and thereby introduce discontinuities or incompatibilities into the survey. Remember: the overriding objective is to stay as close to the original plan as is feasible.

8. Editing and Processing Controls

8.1 Introduction. When editing and data processing begin, almost all that can be done in planning a good survey and in securing the basic data input has been done. The raw material has all been assembled. The task is now to edit, refine, process this raw material into clean, coded, properly classified, numerical and categorical measures — and to do this with a minimum of error.

This is an easier task than was the collection of data, and the establishment and maintenance of controls is easier. Control procedures are more straightforward because the operations that must be performed in editing and processing are more clearly definable, it is easier to tell what is right and what is wrong, and the operations have the character of standardized mass production, which is more readily subjected to uniform surveillance than are the highly variable interviewing situations.

Needed control devices fall into four classes, each of which will be discussed separately.

A. Document flow control
B. General editing procedures
C. Multiple-coding of subsamples
D. Transcription control

8.2 Document Flow Control. This is perhaps the simplest control phase in the entire survey operation. Yet, too often it is done in a sloppy manner, if at all. Unless care is taken, gross misadventures can occur, and they are inexcusable. The objective is simple. It is to assure that all the data and evidence which have been collected and should go into tabulation and estimation are in fact processed, properly labeled, and accounted for, and that none of the data are lost.

The first requirement is a plan for processing: a flow chart that tells what documents move from what point to which other point, and what operations should be performed on the documents at each point. In order to keep track of questionnaires and related papers as they move through these channels, it is almost essential that each sheet of paper have an identifying code and serial number. Control counts of the various types of sheets must be established at points as close as possible to the time at which the sheets are initially posted or executed. As the papers move along the production line, these control counts must be verified frequently to insure that all papers which started a journey are still present.

This bulk control is facilitated if documents are processed and transmitted in blocks or "batches." The documents in a batch can be physically banded together in a file-folder or envelope along with a transmittal sheet which identifies the documents, their batch number, the number of documents in the batch, dates of transmission from specified points, initials of the clerk who is responsible for verifying the number of documents, and perhaps carrying provision for certification by other clerks that particular processing operations have been performed on the documents in the batch. A space for comments will be provided on the transmittal sheet.

A supplementary procedure also is necessary. This is a "safety-valve" provision. Clearly, it is not feasible to stop all production and search indefinitely for a document which the controls discover to be missing. It might take a long time to re-locate the missing item; it is possible the document has been destroyed and can never be found. The supplementary procedure has three aspects or phases.

1. Initiation of a "Supplemental Batch No. ____" which contains a dummy of the

missing item(s), describes the problem, and is keyed to its parent batch through identifying codes. The supplemental batch is then put on a separate track in which the missing information is searched for, while the parent batch moves through regular processing. Later, the supplementary batch is processed specially until it "catches up" with its parent batch, if the missing material is found, or a decision is made that the missing material cannot be found. (Of course no supplemental batch is created until a modest effort has been put forth to make the parent batch intact without undue delay in normal processing.)

2. The parent batch transmittal sheet is annotated to show that the supplemental batch has been initiated and to permit proper bookkeeping at subsequent control points.

3. At appropriate later points, the data or evidence from the supplemental batch is integrated with that in the parent batch. The detailed steps in integration will depend on the nature of the missing data, whether the missing material was found, and on the timing of the "catch up." It may be possible to re-insert the located missing material in the parent batch and proceed from that point as though it had never been lost. Or it may be necessary to adjust later at some stage of tabulation or estimation. (Should the latter occur, proper action is covered under *Imputation* in section 9.)

8.3 General Editing Procedures. Central editing of data reported from the field has two general purposes: (1) To minimize the differences of reporting on individual questionnaires to the extent that care and reasonable judgments based on the evidence at hand make possible, and thus to submit to tabulation data that are as nearly accurate as circumstances at this stage permit. (2) To maintain running tabulations of numbers of errors discovered, by type, and by whom made, and to feed these results back to the field, so that supervisors can be guided in reducing mistakes in later cycles of data collection.

The mechanics for accomplishing these purposes will vary with the precise content of questionnaires, with available data-processing equipment, and to some degree with experience and the availability of collateral data which may be useful for checking such things as acceptable magnitude or range of entries for particular items. The principles, guidelines, and illustrations offered in subsequent paragraphs of this section should be directly applicable in many situations, and suggestive of appropriate steps in others. (Formal quality control over coding will be treated in section 8.4.)

A written loose-leaf manual of editing instructions should be prepared. First drafts of this manual should have been developed in the pre-test and pilot stages of the survey. The manual should cover all aspects of editing. It is best put into loose-leaf format because additions and changes will be found desirable as experience accumulates, and new situations are encountered. For later reference, each page should carry an effective date. An index to the manual will be a great convenience. For ease in handling, it may be good to assemble the manual into more than one book or volume. The manual should be quite specific and explicit: It should direct the clerk as to exactly what to do in each of the commonly encountered situations.

A general strategy for editing needs be adopted. Three broad classes of faulty data will be found, and remedial action selected for each.

A. Omissions — there is no entry for an item for which there should be an entry.

B. Incomplete entries — part of the entry is present but part is not, e.g., the name of a village is entered, but its district is not shown.

C. Incorrect entries — evidence either internal to the questionnaire or external to it indicates that the entry cannot be correct, or is extremely unlikely.

Decisions must be made for each item or class of items whether the clerk is expected to supply missing data, and if so from where to obtain the needed information; and whether and how to attempt correction of an entry judged to be incorrect.

For some items, a missing or incomplete entry can be supplied by the clerk with little chance for mistake. Examples are cycle number; a code for an interviewer (determined by handwriting, or location of respondent, or ini-

tials on a transmittal sheet); sex of respondent (from respondent name, or content of questionnaire); number of persons in the household (by counting names of members); the district number of a named village (by reference to a catalog); or number of nights spent in a hospital (if admission date and discharge date are shown). For a few missing items, it may be feasible to secure the needed information from a supervisor, or even from the field — but usually this will entail operational delays which cannot be afforded. There are also certain types of corrections which the clerk can make with low possibility of mistake. First, are the cases in which a faulty entry is made for an item for which the clerk could have supplied the data had the entry been blank — cases like those in the examples above. There are others. If date of birth and age are inconsistent with one another, the one with greater probability of error can be adjusted to match the other. (The better choice may vary among different societies. In the U.S.A., date of birth is more likely to be correct.) If the sex had been entered as female, but several other entries obviously relate to a male person, the sex code can be corrected to male. Alternatively if both a male and a female appear on the same sheet, and an obviously female health condition has been credited to the male, it seems reasonably safe to transfer the condition to the female, especially if the physical arrangement of the questionnaire makes a mistake of this type not unlikely. If the year of data reference is, say 1972, but the entry reads "1872," it can be corrected to "1972." If the question was, "How many days last week were you confined to bed?", and the entry reads "30," it is probably good editing to change the entry to "7," on the ground that the "30" refers either to month or to the total episode. In any event "7" reflects a smaller error than "30," and will be less damaging to summary statistics.

Another type of decision is determination of which editing changes should be made by the clerks, and which should be made by machines. In the first instance, a planner should think in terms of the logic of a contemplated editing step. Probably in the majority of circumstances, it will be concluded that a clerk should perform the edit. But if the logic is straightforward, and the number of possible corrective steps not large, and if suitable equipment is available, it may be more efficient, and more readily controllable, to perform the edit with an electronic computer. Examples are given in the following two subsections.

Any simple arithmetic editing step is especially appropriate for machine handling. Suppose, for example, the questionnaire had asked for number of days in the previous year that the person had been in each of three labor force status categories, and one report shows:

Not in the Labor Force	180 days
Employed	160 days
Unemployed	90 days

This total, as reported, is 430 days, but of course there are only 365 days in the year. If there is no other specific evidence on the subject, it may be appropriate in editing to multiply each of the reported day figures by the adjustment factor 365/430 = 0.849, and thus control the sum to 365. The computer can be easily programmed to make this edit with no numerical mistake.

More elaborate computer edits are quite feasible. One notable class is the treatment of "outliers." The frequency of occurrence of most events in any population follows a fairly regular pattern or distribution. For example the number of children ever born to women 20-24 years might be reported in the unedited survey is shown in the table on p. 35.

The reported case of 15 children is an "outlier." It might conceivably be a fact, but it is extremely unlikely and is separated from the rest of the smooth distribution. Even if there is one sample case with 15 children, inflation of that figure to universe proportions would almost surely be misleading. Theory for handling outliers is controversial. One practice, preferred by many, is to replace the outlier with another value drawn randomly from the reported distribution, excluding the outlier. An operational procedure for the present illustration has these steps:

I. Form the cumulative percentage frequency distribution shown in the final column of the table on p. 35.

II. Draw a random number between 1 and 100.

III. Compare this number with the cumulative distribution. Say the number is 63; since it is greater than 52 and not greater than 75, this is a "match" with the class which has 3 ever born children.

IV. Substitute "3" for "15" in the record of the outlier. The computer can perform this edit very rapidly when the test and correction phase is scheduled to follow a preliminary tabulation of reported data. In the illustration the procedure would be applied whenever the reported number of children was greater than 9.

No. of Children	No. of Women Reporting This Number of Children	Adjusted Cumulative Distribution	
		Number	Percent
0	10	10	7
1	20	30	22
2	40	70	52
3	30	100	75
4	15	115	86
5	8	123	92
6	4	127	95
7	4	131	98
8	2	133	99
9	1	134	100
10	0	.	.
11	0	.	.
12	0	.	.
13	0	.	.
14	0	.	.
15	1	.	.

As noted earlier, in clerical operations the clerk should have instructions which tell him what to do in any situation. This means there must be a residual instruction which tells him to "consult supervisor" when no other rule appears in the manual. The supervisor will make an emergency decision and maintain a log of such decisions. When the log indicates any substantial frequency of a particular kind of case, an addition or amendment should be issued for the manual of editing instructions.

Running counts of all editing changes should be maintained, whether the edit action is taken by clerk or by computer. Tabulations of these counts, as noted earlier, are sent to the field for action. They are also made available to analysts for their guidance, and in summary form should be included in published reports as a part of the technical description of the survey.

The clerical edit is a control process. It is a review of operations done by the interviewer. Supervisors should give some attention to whether the editors themselves are performing in a standardized fashion. This is best accomplished through two types of action. The first is a systematic re-edit by the supervisors of a small sample of the work of each editor, and occasional additional spot checks of those who perform less well. The other comes from (probably informal) records of relative frequency of finding errors by the various editors. Wide variations among editors indicates the need for further spot-check re-editing, and perhaps additional training.

8.4 Multiple Coding of Subsamples. Usually such attributes as sex, marital status, religion, race, labor force status, occupation, income, or diagnostic health conditions are reported in terms of words, checked boxes, or ranges of numbers. In order to be tabulated, and analyzed it is desirable or even necessary to convert these terms into coded numbers. Here too the translation needs be done with minimum tolerable error. As with other survey operations, a plan is needed; specifications and instructions prepared; coding clerks selected, and trained; and a control system installed. A detailed account of all the actions that should be taken, the coding guides and indexes written, the forms designed, the training materials prepared, would require much more space than can be afforded here. But it should also be unnecessary for those who have had any experience with coding operations, and who have become acquainted with the spirit of these guidelines.

Nor will there be presented much of the theory that underlies mechanisms which will be recommended, and no attempt will be made to discuss competing or more elaborate approaches than the one suggested here. Suffice it to say that alternative courses do exist, and

they can be studied in, or adapted from, the literature of industrial quality control.

The objective of control over coding is two-fold: first, to assure that the coding process is under control as a process, and second to see that the performance of individual interviewers is adequate and within specified tolerances. We shall present in outline form an illustration of one well-tested and recommended structure for satisfying this objective.

Assume that the task under observation is the assignment of one of 100 possible codes to the health status of each sample person included in the survey. If it is the West Dakota survey, this means 40,000 codes per cycle, or about 160,000 codes over a year of interviewing. Perhaps the reports are being processed in batches of approximately 100 households containing about 400 persons. This would represent a block of 5 consecutive assignments for one interviewer in the field. It might reflect one day's work for a coder. Assume there are 9 coders who do this work, along with other tasks.

The coding control system which is adopted is what is called a 3-way independent sample verification. The operating procedure begins with a batch being assigned to Coder A for coding all 400 sample persons in the batch. (For convenience in discussion it will be assumed that all batches are of the same size, although in practice they will vary somewhat.) This same batch is later assigned to Coder B, with instruction to code a predesignated 10 per cent of the cases — the mechanics of form construction and posting of codes being such that B does not know codes A assigned. Then the batch is assigned to a third Coder C, who, independently of A and B assignments, also codes the same predesignated 10 per cent subsample. The three codes for each of the subsample cases are compared — either manually or by machine. In normal operation, the code assigned by A is the one that enters into tabulation, of which more later. But the comparison becomes the basis for subsequent control decisions and further action. Errors are assigned to coders in accordance with these rules:

Rule I. If A, B and C all agree, no coder is charged with an error.

Rule II. If two of A, B, C agree, but the third does not, the third one is assigned an error.

Rule III. If all three A, B, C have different codes, each is assigned an error.

Data from the operation just described permit the calculation of error rates for each coder. Let y_{ij} be the number of codes assigned by coder i in batch j, and x_{ij} be the number of his errors in that batch. His error rate for that batch is $R_{ij} = x_{ij}/y_{ij}$. By summing over batches, an error rate for a coder over a larger trial period can be secured:

$$R_i = \frac{\sum_j x_{ij}}{\sum_j y_{ij}}$$

By summing over both coders and batches, an overall process average error rate is obtained:

$$R = \frac{\sum_i \sum_j x_{ij}}{\sum_i \sum_j y_{ij}}$$

The observed values R_{ij}, R_i, and R are of course estimates of the true underlying error rates, being based on samples of n_{ij}, n_i, and n observations respectively. Pilot studies and survey requirements will have established a maximum target overall error rate $\overline{R}$ which is both feasible and acceptable. Training must have been continued until in a trial period the observed error rate R meets or betters the target $\overline{R}$. The problem then becomes one of keeping the process in control at that level.

More precise and elaborate models could be constructed, but a simple binomial model is adequate for building a control mechanism. This means that analysis of the observed values R_{ij}, R_i, and R will be carried out under the assumption that they will be distributed normally about their expected mean values with respective variances of

$$S^2_{R_{ij}} = \frac{R_{ij}(1 - R_{ij})}{n_{ij}},$$

$$S^2_{R_i} = \frac{R_i(1 - R_i)}{n_i}, \text{ and}$$

$$S^2_R = \frac{R(1 - R)}{n}.$$

R_{ij}, R_i, and R are all expressed as relative numbers less than unity, rather than in percentage terms.

As was the case in controlling the interviewing process, discussed earlier two types of analysis will be useful One builds on the concept of "acceptance sampling," and the other utilizes the "control chart technique."

Consider first the procedure for accepting or rejecting the coding on a particular batch. Assume that the target acceptable error rate $\bar{R}$ is 10 per cent. (In language previously employed this is the AQL value.) Build a table like Table 8-1, which shows the maximum number of errors, as detected by the 3-way certification procedure, which will permit the retention of the batch as originally coded. If this number of errors is exceeded, the entire batch of 400 cases (not just the 10 per cent sample of 40 triple-coded cards) must be recoded (dependently) by another coder who has been performing at an acceptable error level. (This procedure makes the assumption that the measured error rate in the 3-way verification applies equally to coders A, B, and C, and in particular to production coder A. For the immediate purpose, this is a justifiable assumption.)

Table 8-1 Acceptable Batch Error Rates for Coding of Health Conditions

Number of Conditions Checked in Subsample n_{ij}	Acceptable Number of Errors in the Batch Subsample c	Approximate Value of $P_{0.10}(100)$
25-34	5	30%
35-44	7	29%
45-54	9	28%

Thus, according to the table, if there are 40 conditions coded in a batch subsample, and 7 or fewer errors are found in the 3-way independent verification, the batch is accepted, and moves on its production route. If 8 or more errors are found, all 400 cases must be recoded, dependently. The control process assumes that if the true error rate is 10 per cent or less, the batch will be accepted 95 per cent of the time, while there is only a 10 per cent chance $(P_{0.10})$ that a batch containing an error rate of 29 per cent or greater will be accepted. Reference [7] or other sets of quality control tables can be consulted, if the survey managers should wish to set different levels of control.

The control chart technique may be used in surveillance of the performance of individual coders or of the overall coding operation. The procedure is similar to that described on page 29 for the establishment of a control chart. It is illustrated now for control of an individual coder. The central horizontal line in the control chart, Figure 8-1, is the target maximum permissible error rate $\bar{R}$. The plotted points are values R_{ij} for the i^{th} coder for successive batches j coded by him either as the production coder or as the "verifier." The standard error bands are located through computation of *average* variances, using the formula

$$S^2_{R_{ij}} = \frac{R(1-R)}{n_{ij}}$$

So long as the coder stays within the area bounded by the lines at $\bar{R} \pm 2S_{R_{ij}}$, he is "in control." Whenever a plotted point falls outside this band, the work of that coder should be given a general review by the supervisor. If a second point falls above the band, the coder should be withdrawn from the production process and either retrained or replaced. If a majority of points fall below the central line R, and in particular when points fall below the line $R - 2s_{R_{ij}}$, the i^{th} coder is performing at a lower error rate than the required target. Several interpretations are possible. First, it may be expected that some coders will do better than others; there will be a dispersion of coder averages about the central $\bar{R}$, and this is entirely appropriate. It may be that the coding process on the whole is proceeding at a better-than-expected level of accuracy. That is probably good, and may indicate that a new value of $\bar{R}$ should be adopted in subsequent analysis. Whether this is the case can be checked by combining evidence from all coders, as outlined in the next paragraph. It could be that Coder i is spending more time per case than was allotted for the coding operation — and more than is necessary — and obtaining less

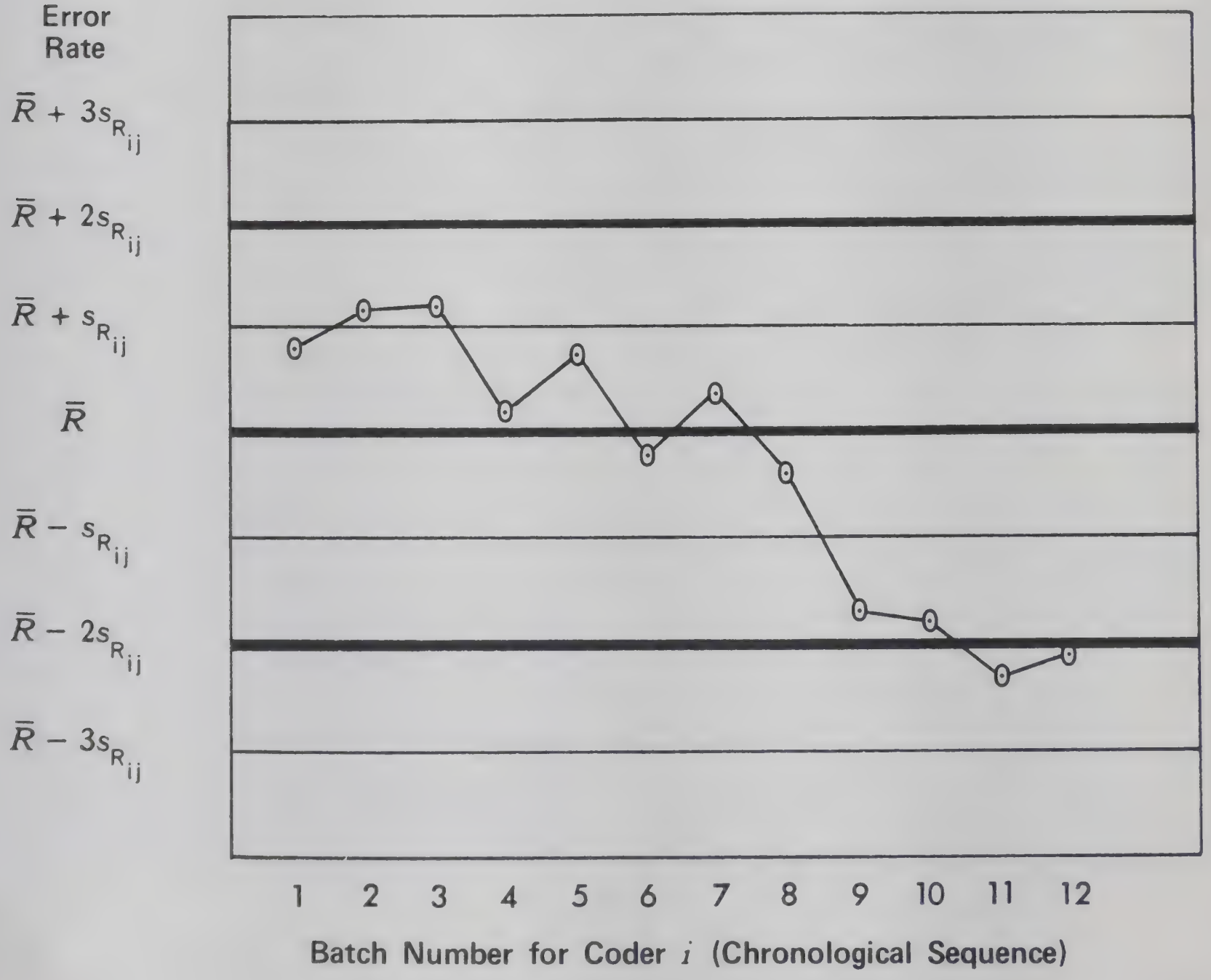
production than is desired. The supervisor should be able to determine if that is so, and to act accordingly.

Surveillance of the overall coding process can be maintained by consolidating evidence from all coders, and over two or more coding items. The process consists of forming weighted error rates and variances, analogous to those described in the section on establishing an appropriate control chart. The plotted points will be the consolidated weighted error rates for a time period of coding — perhaps 5 to 10 working days.

The procedure which has been recommended for controlling the coding of diagnostic conditions is a 3-way independent scheme. It is more elaborate than a simpler 2-way dependent check would be. Dependent checking is adequate and effective for many operations, such for example, as in key punching, or in the reinterview subsample described in section 7. But dependent checking is dangerous in a difficult coding operation, because the second coder is too much influenced by the action of the first coder. Dependent checking lacks the power of true replication in evaluating precision in the assignment of codes.

8.5 Imputation. Early in the present chapter, there was a discussion of several types of editing actions that might be taken to attenuate the damage caused by missing data. Even after editing steps of the kinds treated have been taken, there may well be residual instances in which particular measures or values are missing for some sample cases. The manner of dealing with these residual deficiencies is usually called "imputation." In one sense imputation may be considered a part of data processing, but perhaps more logically it is a component of estimation, and is dealt with in the next section.

Figure 8.1. Control Chart for Coder i on Health Condition


Error Rate
$\bar{R} + 3s_{R_{ij}}$
$\bar{R} + 2s_{R_{ij}}$
$\bar{R} + s_{R_{ij}}$
$\bar{R}$
$\bar{R} - s_{R_{ij}}$
$\bar{R} - 2s_{R_{ij}}$
$\bar{R} - 3s_{R_{ij}}$
1 2 3 4 5 6 7 8 9 10 11 12
Batch Number for Coder i (Chronological Sequence)

9. Control of Tabulation and Estimation

9.1 Character of the Chapter. Except for some remarks on Imputation, this section contains no theory, and few general principles. Rather, it recommends certain precautionary procedures, and offers some hints, designed to reduce the risk of computational boners. Mistakes will occur in transcription and in tabulation. But failure to detect and correct all but a trivial number of such mistakes is inexcusable.

9.2 Transcription Controls. The specific mode of data transcription from original edited documents to tabulation media will vary, of course, with the equipment and processes being utilized for tabulation. Whenever possible, the preferred mode of transcription is an optical, electronic, or mechanical device which is inherently nearly free of error. Often such facilities are not at hand. Then a human transcription must be made by manual copying, keypunching, or some similar operation. In such instances, some form of replicate control is indicated: through either 100 per cent or sample verification. Consider keypunching cards as the transcription process. If the total volume of keypunching is not too great, complete replicate punching may be the best course, using standard keypunch verifiers. If volumes are very large, a sample verification likely is better, using the same kind of plan as was described in section 8 for controlling the coding operation or a dependent verification similar to that described in section 7 for interviewers.

9.3 Control Totals. As a further control on transcription, and over a variety of subsequent tabulating steps, it is highly desirable that a series of "control totals" be established. These start with the case and document counts which were established in the data processing operations. For example, earlier steps in the West Dakota Survey may have established that the survey consists of 40,125 households containing 158,316 persons. Every subsequent tabulation of data which relate to households or persons should carry items of case count which match these controls *exactly.* In any instance in which the counts differ, a reconciliation should be undertaken promptly.

The control total device can be extended beyond the overall case counts. Often it will be feasible to establish other quantitative aggregates in preliminary simple tabulations. For example, the number of males in the sample, the total number of births, or the number of persons in the labor force may be used as subsidiary control totals. It may even be desirable to tabulate a "nonsense" control number, such as the sum of serial numbers of households. The objective again is to have check points for totals which should appear in later tabulations.

The table planning should provide for cross-checks from one table to another. For example, if one table tabulates persons by race and sex, and another by race and occupation, one step toward control should require that a check be made to assure that the race totals, by subclass, agree in the two tables.

9.4 Imputation Procedures. For reasons of convenience in checking against control totals, it is desirable in tabulating to account for every case, and in some manner to construct data for units and items which are still missing after the earlier editing steps have been completed. But there is a more imperative reason for giving careful explicit attention to these imputation procedures. The first point to recognize is that when data are missing, imputation *must* take place, either implicitly or explicitly. If the results of the survey are to be used at all, the analyst or consumer is compelled to draw conclusions about the missing evidence. It is not a question of *whether* to impute, but *how.*

Some analysts might hesitate to "fabricate" data for missing entries, thinking that to do so is cheating. But suppose for a given item, 10 per cent of the entries are missing. If no explicit action is taken, aggregate values from the survey will be of the general order of 10 per cent low, while consumers of the report will be asked to believe and indeed forced to assume that with respect to rates or ratios, the 10 per cent missing cases have the same average value as the overall average of the 90 per cent which were reported. Often this will be an unjustified assumption. Usually a superior imputation procedure is available. The

principle of the superior procedure is to ascribe to non-respondent or missing items the characteristics of respondents who, on the evidence available, are most like the non-respondents. The merit of this principle is evident from the study of a simple model for a survey of a universe consisting of two subclasses of units, the subclasses being internally homogeneous, but with different mean values. The true mean value of a statistic for this universe can be written

$$\bar{x} = X/N = P_1 \bar{x}_1 + P_2 \bar{x}_2$$

where $\bar{x}_i$ is the mean for the i^{th} subclass and P_i is the proportion of universe units in the i^{th} class; i.e., $P_i = N_i/N$, where N is total number of units in the universe, and N_i the number in the i^{th} class. When enumeration is complete, or when 100 per cent response is obtained in a probability survey, the estimate of universe mean is

$$\bar{x}' = P_1 \bar{x}'_1 + P_2 \bar{x}'_2$$

where $\bar{x}_i'$ is an unbiased estimate of $\bar{x}_i$ and the expected value of x' is accordingly x.

Suppose response is incomplete, with $\dot{x}_i$ being the aggregate x reported for the i^{th} subclass by $\dot{n}_i$ units. If no explicit adjustment for nonresponse is made, the estimate of overall mean value will be

$$\bar{x}'' = \frac{\dot{x}_1 + \dot{x}_2}{\dot{n}_1 + \dot{n}_2} = \frac{\dot{x}_1 + \dot{x}_2}{\dot{n}}$$

It can be assumed that $\dot{x}_i/\dot{n}_i$ is a nearly unbiased estimate of $\bar{x}_i$, since the subclasses are stated to be homogeneous. Thus the estimate $\bar{x}''$ can be written

$$\bar{x}'' \approx \frac{\dot{n}_1}{\dot{n}} \bar{x}'_1 + \frac{\dot{n}_2}{\dot{n}} \bar{x}'_2$$

The expected value of $\bar{x}''$ will not in general be equal to $\bar{x}$ unless the nonresponse rates in the subclasses are equal. But this situation can be improved by adopting the recommended principle of imputation: all data in the i^{th} subclass are multiplied by the adjustment factor $n_i/\dot{n}_i$, where n_i is the designed number of units in the survey or census. When this is done the adjusted estimator becomes:

$$\bar{x}''' = \frac{\dot{x}_1 \frac{n_1}{\dot{n}_1} + \dot{x}_2 \frac{n_2}{\dot{n}_2}}{\dot{n}_1 \frac{n_1}{\dot{n}_1} + \dot{n}_2 \frac{n_2}{\dot{n}_2}} = P_1 x_1' + P_2 x_2'$$

and once again the expected value of the estimator approximates the true value. The general principle can be extended, of course, to any number of subclasses.

How should the subclasses within which imputation is carried out be chosen? Three rules, taken together, provide an answer to this question.

A. Choose subclasses for which the difference between means $\bar{x}_1$ and $\bar{x}_2$ is, or is likely to be, substantial.

B. Choose subclasses for which non-response rates are different.

C. Choose subclasses which are of sufficient size (n_i) that they have more than trivial impact on overall results, and at the same time assure that there will be enough responses in the subclass to yield a stable estimate of the imputed values.

Still another reason for not using very small cells for imputation is that one does not wish any adjustment factor $n_i/\dot{n}_i$ to be greater than 2.0. If a factor greater than 2.0 would be required for some cell, it is better to consolidate the cell with a "neighboring" cell, so that the adjustment for the two (or more) combined cells is less than 2.0. Good boundaries for the imputation subclasses will be established in terms of age and sex of sample persons, and of geographic proximity of sample units, since persons of similar age, sex, and geographic location are likely to be more homogeneous with respect to many characteristics, than will persons in a more general population.

For use by analysts, and by readers of the final report, a record should be built and maintained, showing the numbers of imputations made for each item or class of items.

9.5 Validation of Programming. Whether tabulation is accomplished by desk calculator,

punched card electromechanical equipment, or computers, there must be sets of rules, instructions, or "programs" to direct that tabulation. In a major survey, these directions may be quite extensive and elaborate. Whether the directions are expressed as instructions to clerks or as computer programs, they may contain flaws, either because they include incorrect statements or because they fail to provide for certain contingencies that may arise. One technique which will reduce the chance of such error is the use of "hot deck" dummy processing. In this procedure, the statistician designs and then creates a modestly sized set of hypothetical observations similar to those he expects to find in the real survey. Using this set of data, or "hot deck," and with very careful supervision of expert technicians, he carries through manually a complete sequence of estimation and tabulation of a number of key statistics. These carefully determined values become controls. Then the "hot deck" is processed through the contemplated regular production channels and results compared with the controls. If the results and controls agree, the programming is considered validated. If there is disagreement, the source of the trouble is located, and corrected.

We speak of "designing" a hot deck. There is an art in this task. A good design will seek to incorporate a considerable range of types of cases, and at least some of the most likely unusual items of data, in order to provide fruitful tests of the adaptability of the programming to uncommon as well as to more typical observations. The hot deck should contain such items as the following, to make sure programming provides for them:

A. Blank entries where a number is expected.

B. Zero entries in locations in which the normal entry is the denominator of a fraction.

C. A number with more digits than may have been allowed for the field.

D. Numbers which, when multiplied together, will yield an unusually large product.

E. More complex combinations, e.g., a subtraction which is followed by extraction of the square root of the difference — with the hot deck data producing a negative difference.

9.6 Special Gross-Tabulation Checks. There is little point in complete re-tabulation of survey data, especially if the tabulation is accomplished with electromechanical or electronic equipment, since with only minor exceptions the re-tabulation will be an exact copy of the original, and would provide no control. But there are at least two ways in which the general concept of replication can still be useful in tabulation:

A. If certain control totals are established by running an adding machine tape, these should be checked by a replication of the operation before they are accepted. Other manual computational operations which may be a part of the data input, or a summary of output, or an integral part of data programming should likewise be verified through repetition of the operation by a second person.

B. Perhaps the full survey encompasses an extensive record for each of 100,000 or more persons. A tabulation plan which breaks the overall record set into subsets, each of which is processed separately, has several advantages. Results from an early subset may indicate a processing deficiency which can be corrected before the entire survey has been erroneously processed. Results from different subsets can be compared with one another, with the degree of consistency serving as an indicator that tabulation is proceeding well, or perhaps that difficulties are being encountered. The subsets can be formed in many ways. It may be possible to tabulate in randomly formed subsets, which have the additional advantage of facilitating variance calculations. But regional subsets, or demographically-bounded domains, or parts of the long record for each person or unit can serve usefully.

9.7 Comparative Analysis. Are the tabulated results reasonable? This is an imprecise question. It likely has no uniquely correct answer. Opinions will differ, and circumstances will have an impact on how one should search for an answer. In any event, a statistician should always subject his findings to this test: Are these results generally consistent with other evidence and knowledge which I possess, or are they contrary to expectations?

There usually will be two broad classes of other evidence — that which is internal to the survey, and that which is external. The former is always present in considerable degree; the latter will vary rather markedly from one occasion to another. In both cases, the pattern of investigation is what may be called "comparative analysis": evaluation following comparison of a tabulated figure with what might be "reasonably expected." In the internal analysis, the comparisons are largely concerned with consistency relationships among subclasses and between related statistics. Judgments will be subjective; they will be based on assumptions such as: (1) taking sampling variation into account, frequency distributions and trends should tend to be "smooth" unless there is a positive reason for discontinuity; (2) variables which on an *a priori* basis are positively correlated, should not vary in an opposite manner in the survey; (3) many rates and ratios should tend to vary within narrow ranges among subclasses — more erratic observed variation is a sign of trouble.

The possible external comparisons will depend on the existence of other data which should show similar values. These other data may come from earlier surveys of the same type as the present one; from completely different kinds of measurements; from purely hypothetical or theoretical considerations; from the opinions of experts; or even from "common knowledge."

It will be readily understood that comparative analysis does not require necessarily a close match between survey results and the other evidence. The other evidence may be outdated, weak, or wrong. Indeed, if the survey was justified and well-conducted, the evidence from it likely is the best that can be found. But the comparisons should be made. For in some instances they may uncover flaws in tabulation or estimation that can be corrected, or at least recognized and treated in subsequent analysis. More often, they will focus attention on new findings, contrasts, contradictions, and highlights for which consumers may expect some explanation, and on which final reports should comment.

10. Control Over Report Manuscript

10.1 Analytic Methods. Just as choice of formal design of the survey was not within the scope of this manual, neither is the choice of methods of analysis. The theory of statistical analysis is treated in many excellent textbooks and journals. Here we offer only one watchword. Much of classical analytic theory is derived from models that assume simple random sampling from known distributions, with each observation being independent of the others and free from measurement error. These conditions are rarely satisfied in social surveys. The analyst is therefore well advised to consider critically any specific analytic technique before he applies it to his data, in order to assure himself that it is tolerably appropriate to the problem at hand.

10.2 Transcription Verification. All transcription of data, codes, and labels from basic tabulations to worksheets, from one worksheet to another, and to final report copy, should be checked on a 100 per cent basis by someone other than the person who did the original posting, or by one person reading from the source document, and a second person inspecting the posted information.

10.3 Verification of Computations. Similarly, every computation beyond the raw tabulations should be performed twice, the two calculations being done by different persons. Any discrepancies call for a third computation and a complete reconciliation. In addition, a variety of additional numerical checks are desirable. Computer calculations need not be so checked, after the programming has been verified.

One valuable procedure is a series of recalculations in approximate terms, performed by a professional statistician, to confirm the scale and dimension of final results. The three following examples are illustrative and suggestive.

A. The initial computation was

$$85.14\left(\frac{1752.05 - 10.16}{15.15}\right) = 9789.08.$$

The check,

$$\frac{(8.5)\ 10\ (1.74)\ 10^3}{(1.5)\ 10} = (9.9)\ 10^3 = 9900,$$

confirms the scale of the original result.

B. The initial computation was

$$12.5^2 + 11.9^2 + 14.2^2 + 12.9^2 + 13.1^2 = 8375.$$

The check,

$$5(13^2) = 845,$$

shows that the decimal point was incorrectly placed in the initial computation.

C. The original calculation showed:

Cycle	Average Number of Days of Restricted Activity per Person
First 3 months	12.1
2nd 3 months	11.2
3rd 3 months	11.9
4th 3 months	14.1

The preliminary draft of the report summed these "quarterly" figures and estimated 49.3 days of restricted activity per person per year. The professional checker recognized immediately that the clerks failed to realize that each of the cycle or "quarterly" estimates had already been inflated to an annual level, and that the correct figure should be the average of the four, or 12.3 days per person per year, rather than 49.3.

When a column or row of numbers has been converted to a percentage distribution through division of each number by the sum of the column or row, the divisions can be checked by adding the percentages to assure that they sum to 100. In some cases, it may be desirable to adopt a special rounding rule to bring the total to 100.

It is often helpful to utilize an algebraic identity in order to set up a computation pattern which contains a seif-checking feature. Noting, for example, that

$$\sum_{i=1}^{N} (x_i + 1)^2 = \Sigma x_i^2 + 2\Sigma x_i + N,$$

one can set up a work sheet in this fashion:

i	x_i	x_i^2	$(x_i + 1)$	$(x_i + 1)^2$
1	1.2	1.44	2.2	4.84
2	2.5	6.25	3.5	12.25
3				
.				
.				
.				
.				
Totals	**I**	**II**	**III**	**IV**

Then, if the roman numerals represent the arithmetic sums of their respective columns, the check becomes: Does *IV* equal *II* plus 2 times *I* plus *N*? (*N* is the total number of observations.)

10.4. Completeness of Manuscript. Remember that many readers of the final report will have no evidence on the subject other than that contained in the report itself — and even those readers who may be better equipped, will find it convenient to see the full story of the survey assembled in the one package of the report. This means that the manuscript should be inspected to discover if it does include all classes of material which the reader may reasonably wish to have. This inspection should include at least these check points:

A. In introductory section that presents the framework within which the survey was conducted, and the background for the undertaking. Sponsors and operators of the survey should be identified.

B. Clear statement of the coverage and purpose of the survey.

C. Both a condensed text statement on the design of the survey, and a more detailed account (probably in a technical appendix) which provides a rather full description of the entire undertaking.

D. A Summary of findings and conclusions.

E. Other significant analytic results, clearly exhibited.

F. Detail appendix tables, which will permit a reader to confirm quantitative declarations.

G. Glossary of terms and notations.

H. A section of the report should treat limitations and qualifications of the data; tables or charts of sampling errors should be included;

attention should be directed to possible sources of measurement error, and quantitative evidence on this aspect offered when known.

I. Table of Contents.

J. List of references, and proper keying to those references.

K. If geographic coverage or sub-universes are involved, the inclusion of one or more maps is advisable.

10.5 Proofreading. As many research workers have learned, and all people in the graphic arts and publications industries know, it is nearly impossible to convert an extensive manuscript to printed format without error. This is particularly true when the manuscript includes equations, charts, and many figures. It is, neverthelsss, worth substantial effort to keep errors at this final step to a minimum. When one has devoted much thought and energy to operational control of the intrinsically much more difficult planning and production phases of the survey, care in proofreading the final report seems such an obvious requirement, that it is almost presumptuous to mention it. Mention is made here because the apparent simplicity of the proofreading question often leads to an expensive carelessness.

It is suggested that in addition to an initial "proofing" of final reproduction copy by the typist, compositor, draftsman, or proofreader, there be at least three further proofreading operations:

A. A two-person verification by professional personnel, one reading from manuscript, the other confirming the reproduction copy. This step should include a special operation of verifying legend and posting of any charts.

B. An independent reading of the reproduction copy by a statistician who was not an author of the report.

C. A final critical reading of the reproduction copy by the principal author of the report.

11. Concluding Comment

Some readers of this manual may despair. They may feel the task of control is so extensive and so difficult, and their available resources so limited, that the task is hopeless. There may be other readers who feel that the author is overly pessimistic about performance of survey personnel, and over cautious; that many of the recommended control procedures are unnecessary, or even silly. To the first of these readers, we admit that the sources of possible error are many and diverse, and some of the control procedures are difficult to carry out. But be encouraged. Once instituted, most of the control techniques are less complex operations than they appear at first. To those who think the proposed controls are over ambitious, we observe that every one of the hazards discussed does exist, and suggest that a program of too much control is less dangerous than one of too little control.

To all readers it is emphasized that the primary message is to encourage the adoption of a policy of control, a habit of thinking that recognizes risk of error in any survey procedure, and seeks means of keeping that risk within tolerable bounds. It is not expected that every control procedure proposed will be introduced into every survey. Indeed it would be less than honest to convey the impression that most of the procedures discussed are present in most surveys. Some surveys must be conducted under unfavorable conditions, and with far from ideal personnel and resources. Nevertheless, one of the most serious flaws in survey work is the discrepancy between intention and execution. Controls designed to narrow that discrepancy are among the most profitable investments that a survey manager can make.

Appendix A. A Brief Note on Acceptance Sampling and Process Control

A1. For an account of the theory underlying acceptance sampling and process control, one should consult a textbook on the subject, such as reference [4] or [7]. But it may be useful to note here a few of the key concepts and a little of the terminology in common use.

A2. *Operating Characteristic (OC) Curve.* This curve is the key description of an acceptance sampling plan. It gives the probability of accepting, A, a lot, under a specified plan for any fraction defective, p. The equation for the curve is

$$A \approx \sum_{d=0}^{c} \binom{n}{d} p^d (1-p)^{n-d}$$

where n is the sample size (no. of items tested), and

d is the number of items in the sample found defective,

c is the acceptance number, which means that if $d \leq c$, the lot is accepted, and otherwise the lot is rejected.

The OC equation is usually plotted as a continuous curve (see Figure A-1). The shaded area at the left of the curve shows the region of acceptance for high quality production — where the probability of acceptance is near unity. The shaded area at the right shows the region of acceptance for very low quality work — where the probability of acceptance approaches zero. The shape of the OC curve is, of course, a function of all the parameters of the curve: p, n, and c. In particular, the OC curve is steeper, and thus more discriminating, for larger sizes n. A high level of quality is assured when c is small compared with n, but costs increase as c is decreased for given n.

It will be observed that A is given by the sum of the first $(c + 1)$ terms of the binomial expansion $[p + (1 - p)]^n$. For the smaller values of p (which are the most important, in many situations), the value of A can be approximated by

$$A' \approx \sum_{d=0}^{c} e^{-np} (np)^d / d!.$$

A3. Average Outgoing Quality Curve (AOQ). This curve gives the expected average outgoing percent defective, B, for the chosen plan, when the incoming fraction defective is p. The AOQ curve has the equation

$$B = pA$$

where A is given by the OC curve, and typically has a shape as in Figure A-2.
It will be observed that if the actual incoming fraction is either very low or very high, the expected outgoing fraction defective is low — low because either p is itself low or because the probability of acceptance is low if p is high. The worst possible result is measured by the maximum value of the AOQ curve, called the average outgoing quality limit, or AOQL.

A4. Choice of Acceptance Plans. Theory for construction of ideal acceptance plans is outside the scope of this manual. An outline of one simple approach and approximating method for choosing the acceptance number c makes use of the concept of average quality level (AQL), and can be presented here. One selects a target acceptable error rate (per cent defective), P, as the AQL. Then for each of several likely sizes of sample observations, n, which will have been largely determined by data

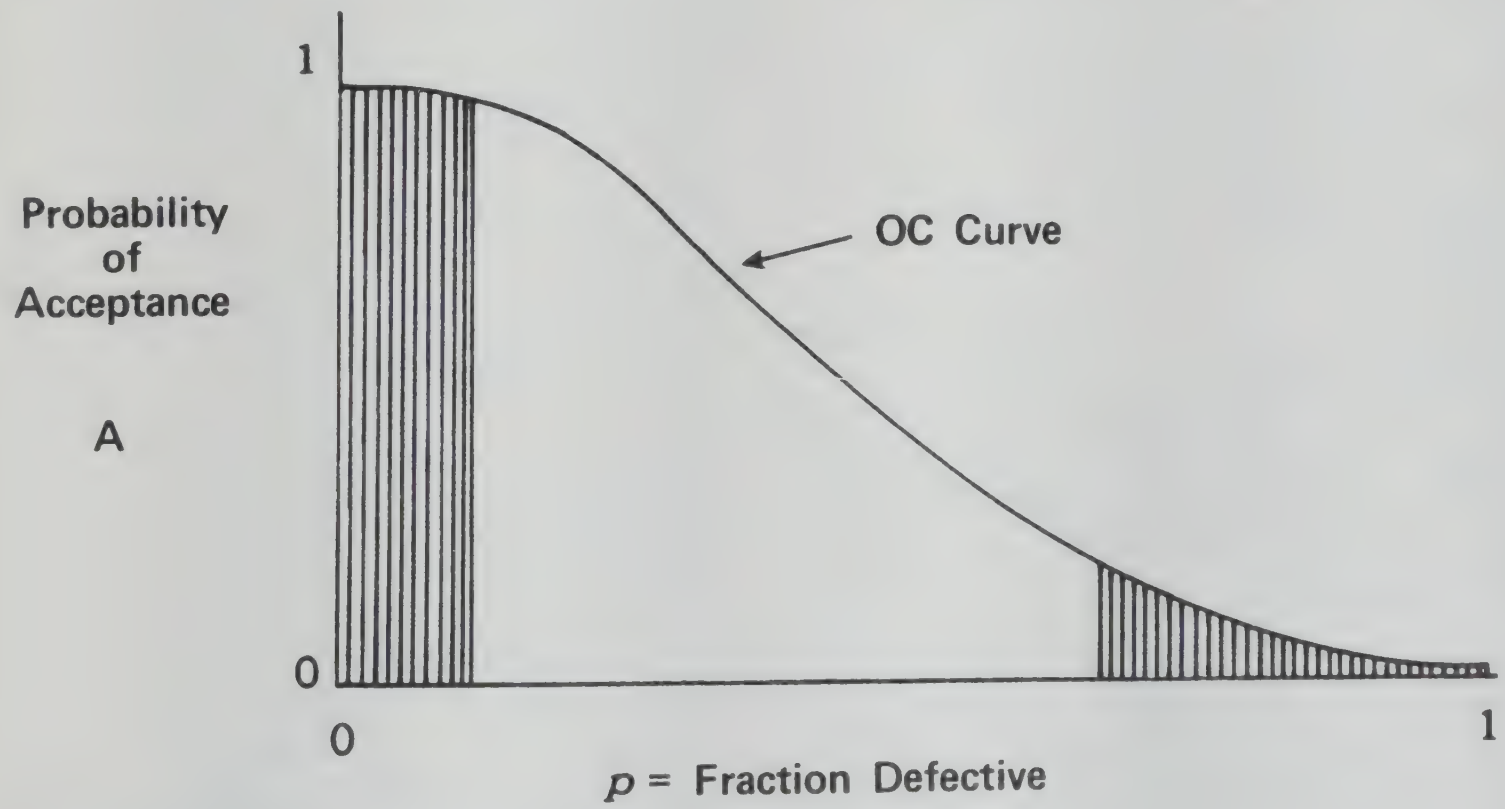


Figure A-1. Operating Characteristic Curve (OC)

collection or processing procedures, a series of calculations is made. Let *D* be a hypothetical number of defectives in *n* observations, where the probability of an observation being defective is *P*, the chosen AQL. The probability of *D* being realized is

$$p(D) = \binom{n}{D} P^{D} (1 - P)^{n-D}.$$

Values of *p(D)* are calculated for $D = 0, 1, 2, 3 \ldots\ldots$ and their cumulative sum obtained until that cumulative sum first exceeds 0.95. If that value of *D* is $(c + 1)$, then *c* is the acceptance number. The rationale is to make the odds 19 to 1 that the lot will be accepted when the per cent defective is *P*. If the actual per cent defective is less than *P*, the probability of acceptance is greater than 0.95. If the actual per cent defective is greater than P, there is a less than five per cent chance of acceptance. This probability can be calculated by the same technique for any other hypothetical value of *P'* different from *P*, by substituting *P'* for *P* in the formulas. A cumulative average observed sample error rate should be computed, and if it exceeds *P* by much, the acceptance number should be lowered, or a more sophisticated analysis undertaken, turning to the basic principles stated in A2 or A3, or to textbooks on the subject.

The final column of Tables 7-2 and 8-1 is concerned with "Consumers Risk." The initial objective was to be 95% sure of accepting the lot if the actual error rate is *P*. The "consumer" wants to know how large an error rate he may be faced with at the 10 per cent acceptance point on the OC curve. This is the figure in the final columns of Tables 7-2 and 8-1. Knowing the sample size, *n*, it can be located approximately by trial and error calculations, or by scanning and interpolating in a table of cumulative sum of terms in a binomial or Poisson distribution. For example, for $n = 40$, and $c = 4$, one finds using the Poisson Distribution, that 0.20 is the value of P' which yields $A' = 0.10$. That is

$$\sum_{d=0}^{4} e^{-40(0.20)} [40(0.20)]^{d} / d! = 0.10.$$

(This same solution can be obtained from tabled data such as is found in Table 6, page 569 of reference [7], where one sees for $c = 4$, and the Poisson summation equals 0.10, that $nP' = 8.0$; and since $n = 40$, P' must equal 0.20.) This means that there is a 10 per cent chance of accepting an error rate of 20 per cent, if one is using an acceptance number of 4 in a sample of 40 observations.

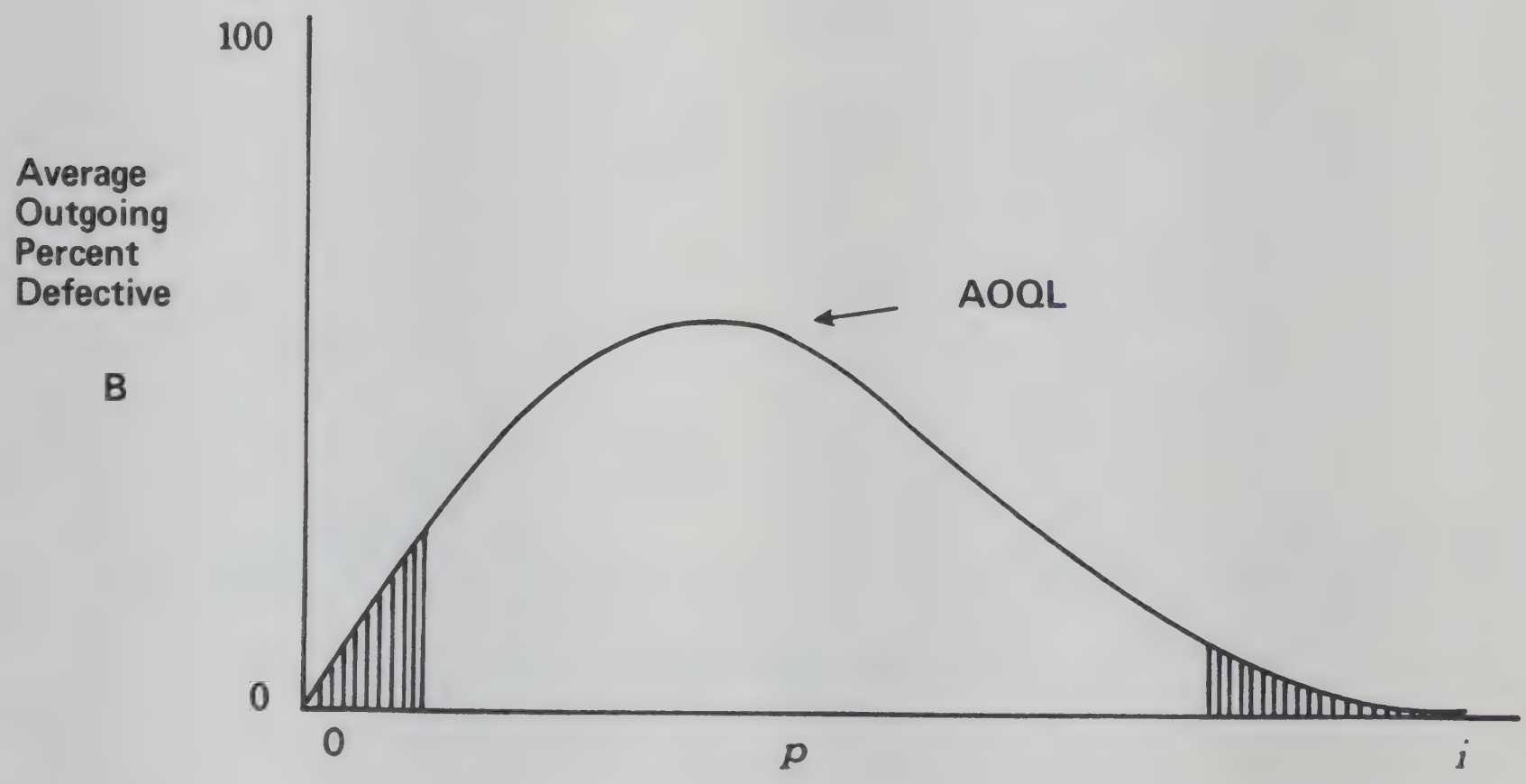


Figure A-2. Average Outgoing Quality Curve (AOQ)

Appendix B. An Illustrative Procedure for Guidance in Selecting and Training Interviewers

B-1. Purpose of the Appendix. Section 7.4 above includes description of a model intended to be helpful to survey managers in setting standards for initial training and selection of interviewers. It was said that the model cannot in practice be followed precisely, but that it might serve as an indicator of how we should proceed. The present index offers an illustration of a pattern of thinking that makes use of the model.

The illustration may appear to exhibit a more precise and sophisticated solution to the problem than is really attained. Although the method of solution borrows from sound statistical theory, the conditions in any given application are such that conclusions must be considered "reasonable judgments" rather than demonstrably validated rules.

The general rationale of a suggested approach will be outlined first, and then presented in somewhat greater illustrative detail. It is recognized that no interviewer will be perfect, and that even those interviewers who are acceptably proficient will vary among themselves in actual performance. The problem is then to determine a reasonable level of *average* performance for the chosen group of interviewers, an acceptable range of deviation from this average for any interviewer who is to be retained, and consequently the error rate for individual interviewers above which the interviewer should be dismissed. An operating solution to the problem is secured through mesh of a theoretical model, speculation about typical performance, and observations in the pretest.

B-2. An Example. The arguments are most easily illustrated in terms of a specific hypothetical example. Perhaps the survey design contemplates for several key statistics a sampling relative standard error of 3 per cent, and has set as working targets measurement relative standard error of 5 per cent [*i.e.*, $V_m^2 = 0.0025$], and relative bias not to exceed 10 per cent. Speculation on the allocation of measurement variance and bias to the various stages of survey operation is a formidable task for which neither theory nor available empirical evidence give much guidance. But it is plausible to argue that at least half the relative bias and a substantial majority of the measurement relvariance will have their source in the agent operations of interview, transcription, and editing-coding. To get at the dimension of the standards problem, let us say that the data collection and processing operations must have a relative bias of not more than 5 per cent, and measurement relvariance of 0.0016. We speculate further that 4 points of the 5-point bias and 0.0009 of the 0.0016 relvariance come from the interview.

Even accepting these speculations, a proper course of action has not yet been determined, for there remain many specific arrangements which would meet the requirements. But useful guidelines are beginning to take shape. Let $\bar{x}$ represent "correct" performance in interviewing. Then the expected value of $\bar{x}'$, the average value of the pooled performance of selected trained interviewers must not deviate from $\bar{x}$ by more than 4 per cent of the $\bar{x}$ value.

A *variance* of the $\bar{x}'$, arising from variation in the performance of individual interviewers, can be allowed up to the amount of $0.0009\bar{x}^2$.

The distribution of individual interviewers by the size of their performance scores is likely to be somewhat skewed, but for the purposes of this exercise, it can be assumed to follow the Gaussian Normal Curve. If the field work calls for, say, 100 interviewers, this means that a population variance among interviewers of $\sigma^2_{\bar{x}_i} = 100\,(0.0009\bar{x}^2) = 0.09\bar{x}^2$ can be tolerated. Thus a relative standard deviation of 30 per cent is permitted.

Consequently, a panel of trained interviewers must be selected for which in relative terms the expected pooled performance is within 4 per cent of the target while variations among the interviewers is held to a coefficient of variation of 30 per cent.

We still do not have an unique solution to setting standards for selection of individual interviewers, but the following plan is a reasonable one, and will provide at least moderate confidence that the interviewing force is capable of adequate performance.

1. Determine the *direction* of expected bias. Usually this will be a negative bias or undercount. That is $E\bar{x}' - \bar{x} = B$ will be a negative number.

2. The model leads, then, to a desired distribution of trained interviewers similar to that depicted in Figure B-1.

3. A key feature of this example is that not more than 15 per cent of interviewers should have "deficiency" scores of more than 34 per cent, nor more than about 15 per cent an "excess" score of more than 26 per cent.

4. A slightly conservative interpretation (appropriate because necessary assumptions have been weakly based) argues that no interviewer should be accepted with a deficiency error as large as 65 per cent or an excess error as large as 55 per cent.

5. The summary guideline then for selecting, training, and rejecting interviewers uses, in this illustration, the following target pattern of test scores.

Criterion	Percent of "Correct" Measure $\bar{x}$
Average pooled error for all interviewers	Not more than 4 percent
Not more than 15 interviewers with a "deficiency" error greater than	34%
Not more than 15 interviewers with an "excess" error greater than	26%
Reject any interviewer with a "deficiency" error greater than	65%
Reject any interviewer with an "excess" error greater than	55%

It should be obvious that the argument of this appendix leads only to rough measures which can serve as guidelines for setting reasonable standards for interviewer performance. The objective has been to determine the general

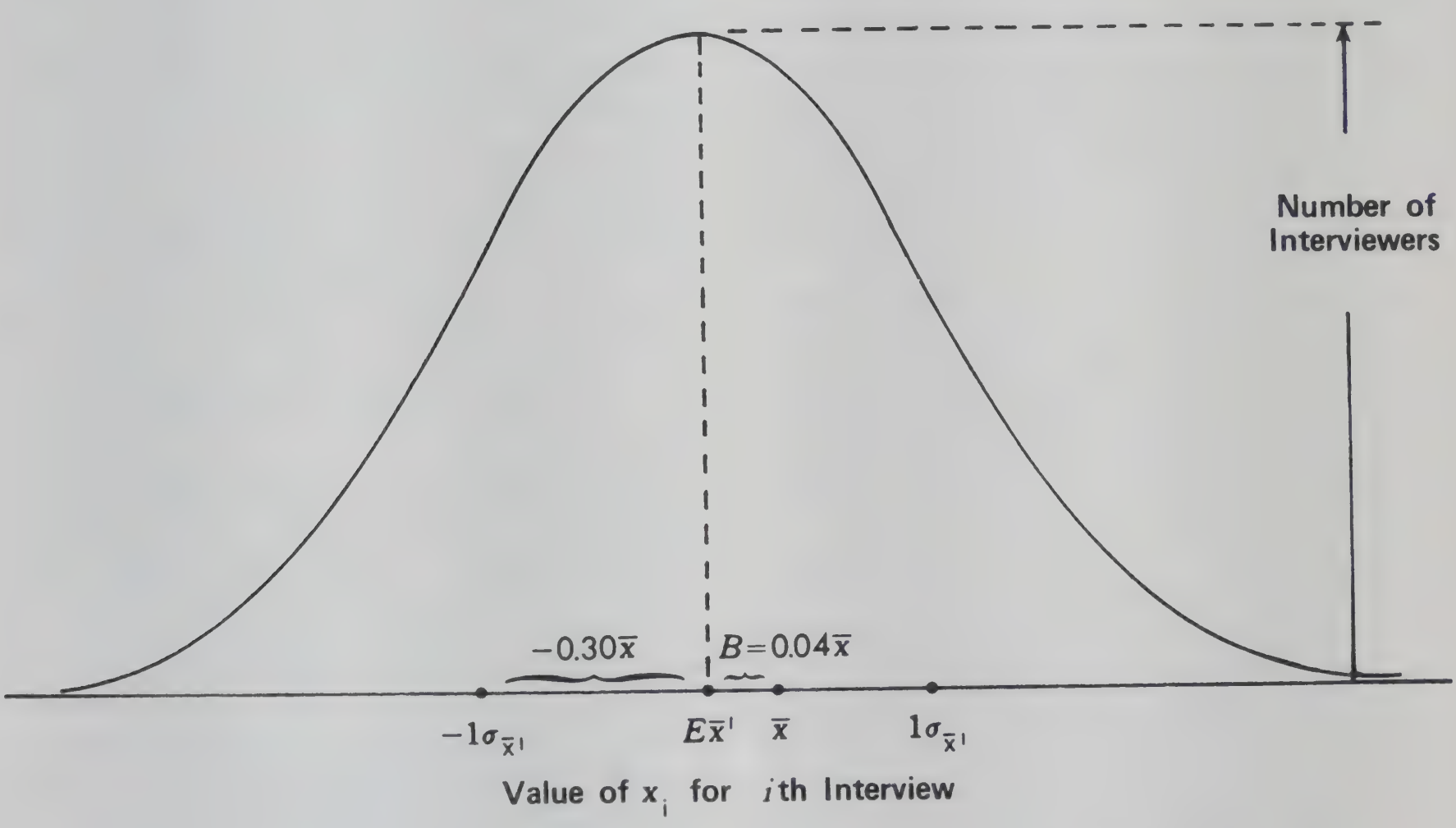


Figure B-1. Expected Performance of Individual Interviewers

dimensions of a rational course. Using a similar pattern of reasoning, and taking account of the peculiar conditions of his survey, the operating statistician can adopt variations of this plan, which should lead to a sensible pattern of training, retraining, rejection of some trainees, and selection of the panel of interviewers.

List of References

1. Chandrasekar, C.; and Deming, W. Edwards, On a Method of Estimating Birth and Death Rates and Extent of Registration, *Journal of the American Statistical Association.* Vol. 44, 1949, pp. 101-115.
2. Cochran, William G., *Sampling Techniques,* 2d ed., John Wiley & Sons, New York, 1963.
3. Deming, W. Edwards; and Geoffrey, Leon, On Sample Inspection in the Processing of Census Returns, *Journal of the American Statistical Association,* vol. 36, no. 215, pp. 351-360.
4. Duncan, Acheson J., *Quality Control and Industrial Statistics,* 3d ed., Richard D. Irwin, Inc., Homewood, Illinois, 1965.
5. Hansen, Morris H.; Hurwitz, William N.; and Madow, William G., *Sample Survey Methods and Theory,* vol. 1, John Wiley & Sons, New York, 1953.
6. Hansen, Morris H.; Hurwitz, William N.; and Pritzaker, L., The Estimation and Interpretation of Gross Differences, and the Simple Response Variance, Processed Paper, U. S. Census Bureau, Washington, D. C., 15 February 1963.
7. Grant, Eugene L., *Statistical Quality Control,* McGraw-Hill, New York, 1964.
8. Jabine, Thomas B.; and Bershad, Max A., Some Comments on the Chandrasekar-Deming Technique for the Measurement of Population Change, (A paper presented at the CENTO Symposium on Demographic Statistics, Karachi, Pakistan, November 1968, prepared in the U. S. Census Bureau, Washington, D. C.)
9. Kish, Leslie, *Survey Sampling,* John Wiley & Sons, New York, 1965.
10. Lauriat, Patience, Field Experience in Estimating Population Growth, *Demography,* vol. 4, no. 1, Chicago, Illinois, 1967.
11. Liberian Department of Planning and Economic Affairs, *Liberian Population Growth Survey Handbook,* Monrovia, Liberia, 1969.
12. Pakistan Institute of Development Economics, Report of the Population Growth Estimate Experiment, Karachi, Pakistan, 1968.
13. Rumford, John C.; Herperkan, Yasar; and Fincancioglu, Nuray, The Principles and Preliminary Results of the Turkish Demographic Surveys, *Public Health Reports,* vol. 83, no. 7, Washington, D. C., July 1968.
14. U. S. Bureau of the Census, The Current Population Survey Reinterview Program, Technical Paper No. 6, Washington, D. C., 1963.
15. U. S. Bureau of the Census, Quality Control of Preparatory Operations, Microfilming, and Coding, Census of Population 1960, Washington, D. C., 1965.
16. U. S. Bureau of the Census, Quality Control of Field Enumeration, Census of Population, 1960, Washington, D. C., 1967.
17. National Health Survey, The Statistical Design of the Health Household Interview Survey, *Health Statistics,* Series A-2, Washington, D. C., 1958.

DEM-130
10984 N72

Publications of the POPLAB Program:

The Scientific Report Series

1. THE CONCEPT AND THE PROGRAM OF THE LABORATORIES FOR POPULATION STATISTICS by Forrest E. Linder
2. ORGANIZATION AND METHODS OF THE DUAL-REPORT SYSTEM IN TURKEY by Eliska Chanlett
3. SUBJECT MATTER COVERAGE IN THE DUAL-REPORT SYSTEMS OF INDIA, PAKISTAN, TURKEY, AND LIBERIA by James R. Abernathy and Anders S. Lunde
4. THE FIRST AND SECOND INTERNATIONAL POPLAB CONFERENCES
5. ORGANIZATION AND METHODS OF THE DUAL-REPORT SYSTEM IN LIBERIA by Joan W. Lingner
6. ORGANIZATION AND METHODS OF THE DUAL-REPORT SYSTEM IN PAKISTAN by James R. Abernathy

The Manual Series

1. MAPPING AND HOUSE NUMBERING by Dorothy S. Cooke
2. OPERATIONAL CONTROL OF SAMPLE SURVEYS by Walt Simmons

The Reprint Series

1. ON A METHOD OF ESTIMATING BIRTH AND DEATH RATES AND THE EXTENT OF REGISTRATION by C. Chandra Sekar and W. Edwards Deming
2. THE USE OF SAMPLING FOR VITAL REGISTRATION AND VITAL STATISTICS by Philip M. Hauser
3. THE DESIGN OF AN EXPERIMENTAL PROCEDURE FOR OBTAINING ACCURATE VITAL STATISTICS by Ansley J. Coale

---

SOME RESULTS FROM ASIAN POPULATION GROWTH STUDIES by William Seltzer